JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

478R

VOLUME 49: 1954
NUMBERS 265-268

Published Quarterly by the
AMERICAN STATISTICAL ASSOCIATION
WASHINGTON, D. C.
1954

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

The Editors welcome the submission of manuscripts for possible publication. They should be typewritten entirely double-spaced, including footnotes, and two copies should be sent to the Editor, W. Allen Wallis, 207 Haskell Hall, University of Chicago, Chicago 37. Books for review should be sent to the same address. Unsolicited book reviews are not accepted, but suggestions of titles for review are welcome.

Errata: Readers and authors are urged to submit to the Editor notices of errors found in this or any previous volume. These will be published once a year, in the December issue.

EDITORIAL COLLABORATORS

Meyer Dwass
Northwestern University
Paul S. Dwyer
University of Michigan
Churchill Eisenhart
National Bureau of Standards
Frederick A. Ekeblad
Northwestern University
Hope T. Eldridge
United Nations
Benjamin Epstein
Wayne University
Helen C. Farnsworth
Stanford University
Robert Ferber
University of Illinois
Enoch B. Ferrell
Bell Telephone Laboratories
Clarence B. Fine
Bureau of Internal Revenue
A. L. Finkner
North Carolina State College
Evelyn Fix
University of California (Berkeley)
John K. Folger
Southern Regional Education Board
Richard J. Foote
Bureau of Agricultural Economics
Robert N. Ford
American Telephone and Telegraph Company
Varden Fuller
University of California (Berkeley)
Alexander Gerschenkron
Harvard University
M. A. Girshick
Stanford University
Leo A. Goodman
University of Chicago
M. J. Gordon
Massachusetts Institute of Technology
Wilson H. Grabill
Bureau of the Census
Howard Whipple Green
Cleveland Health Council
Irving I. Gringorten
Air Force Cambridge Research Laboratory
Emil J. Gumbel
Columbia University
John Gurland
Iowa State College
Louis Guttman
Israel Institute of Applied Social Research
Clauson D. Hadley
University of Santa Clara
Margaret J. Hagood
Bureau of Agricultural Economics
Max Halperin
National Heart Institute
C. Horace Hamilton
North Carolina State College

P. C. Hammer
University of Wisconsin
Morris H. Hansen
Bureau of the Census
Frank Harary
University of Michigan
Arnold Harberger
University of Chicago
Boyd Harshbarger
Virginia Polytechnic Institute
Harry P. Hartkemeier
University of Missouri
H. O. Hartley
Iowa State College
James B. Hassler
University of California (Berkeley)
Millard W. Hastay
National Bureau of Economic Research
Walter A. Hendricks
Department of Agriculture
Clifford Hildreth
North Carolina State College
Jack Hirshleifer
The Rand Corporation
Walter E. Hoadley, Jr.
Armstrong Cork Company
J. L. Hodges, Jr.
University of California (Berkeley)
Wassily Hoeffding
University of North Carolina
Frederick H. Hollander
Institute for Numerical Analysis
Daniel G. Horvitz
University of Pittsburgh
Harry M. Hughes
University of California (Berkeley)
William Hurwitz
Stanford University
Abram J. Jaffe
Columbia University
Julius Jahn
State College of Washington
Rheem F. Jarrett
University of California (Berkeley)
Emil H. Jebe
Iowa State College
T. A. Jeeves
University of California (Berkeley)
R. J. Jessen
Iowa State College
Harold F. Dorn
National Institute of Health
William O. Jones
Stanford University
Norman M. Kaplan
The Rand Corporation
Leo Katz
Michigan State College
Oscar Kempthorne
Iowa State College

J. A. RIGNEY
North Carolina State College
STANLEY REITER
Stanford University
PAUL R. RIDER
Wright-Patterson Air Force Base
HARRY V. ROBERTS
University of Chicago
HENRY G. ROMIG
Pacific Palisades, California
MURRAY ROSENBLATT
University of Chicago
ALAN ROSS
Iowa State College
LESTER SARTORIUS
University of Illinois
FRANKLIN B. SATTERTHWAITE
General Electric Company
LEONARD J. SAVAGE
University of Chicago
HENRY SCHEFFÉ
University of California (Berkeley)
DAVID M. SCHNEIDER
New York Department of Social Welfare
EDWARD E. SCHWARTZ
Children's Bureau
HARRY SCHWARTZ
The New York Times
ELIZABETH L. SCOTT
University of California (Berkeley)
EDWARD S. SHAW
Brookings Institution
G. D. SHELLARD
Metropolitan Life Insurance Company
HENRY S. SHRYOCK, JR.
Bureau of the Census
JACOB S. SIEGEL
Bureau of the Census
ROSEDITH SITGREAVES
Stanford University
JOHN H. SMITH
American University
MILTON SOBEL
Cornell University
HERBERT SOLOMON
Columbia University
ROBERT M. SOLOW
Massachusetts Institute of Technology
MORTIMER SPIEGELMAN
Metropolitan Life Insurance Company
WILLIAM A. SPURR
Stanford University
ROBERT B. STEFFES
Bureau of the Budget
JOSEPH STEINBERG
Bureau of the Census
ALAN STUART
London School of Economics
R. M. SUNDRUM
University of Rangoon
G. H. SYMONDS
Esso Standard Oil Company
CONRAD TAEUBER
Bureau of the Census
D. TEICHROEW
Institute for Numerical Analysis
MILTON E. TERRY
Bell Telephone Laboratories
DOROTHY S. THOMAS
University of Pennsylvania
DONOVAN J. THOMPSON
University of Pittsburgh
G. TINTNER
Iowa State College
C. K. TSAO
Wayne University
JOHN W. TUKEY
Princeton University
DAVID VALINSKY
City College of New York
DAVID VOTAW
Yale University
DAVID L. WALLACE
University of Chicago
JOHN E. WALSH
Naval Ordnance Test Station
ALFRED N. WATSON
Wesleyan University Press
FREDERICK V. WAUGH
Bureau of Agricultural Economics
B. L. WELCH
University of Leeds
ROGER I. WILKINSON
Bell Telephone Laboratories
S. S. WILKS
Princeton University
RAMSAY WOOD
Federal Reserve Board
MAX WOODBURY
University of Pennsylvania
F. YATES
Rothamsted Experiment Station
MARVIN ZELEN
National Bureau of Standards

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

CONTENTS OF VOLUME 49

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 265 MARCH 1954 *Volume 49*

THE PRESENT STRUCTURE OF THE ASSOCIATION*

WILLIAM G. COCHRAN
The Johns Hopkins University

FIVE years ago, the Association adopted a new Constitution which was intended to facilitate substantial changes in the nature of the Association. Written Constitutions are not noted for their ability to grip and hold the reader's interest, and I doubt whether many members paid more attention to the new Constitution than was necessary in deciding how to vote on it in 1948. Consequently, I would like to present some impressions of the experience of the Association during the first five years of operation under the new Constitution. I hope that this account will give members a better picture of the present nature of the Association and will lead up to several questions concerning our future development about which I wish to encourage members to do some thinking.

THE SITUATION AS IT APPEARED IN 1945

Planning for a new Constitution began when the Association was able to resume normal activities towards the end of World War II. In the early discussions about a suitable future pattern for the Association, the committee at work on the new Constitution took note of four developments in the field of statistics that seemed relevant.

1. Statistical techniques had penetrated into a great variety of fields. Up till about 30 years ago, practical statistics dealt mainly with applications to economics, business and government, and the interests of the Association's members tended to reflect this fact. It

* Presidential Address at the 113th Annual Meeting of the American Statistical Association, Washington, D. C., December 28, 1953.

is easy to exaggerate the extent to which this was so: the Association has always welcomed statisticians in any field of knowledge and 30 or 40 years ago the *Journal* was publishing important papers on a wide range of topics. But the organized activities of the Association dealt largely with applications in the economic sphere. In the 30's, however, and still more in the early 40's, the increased use of statistical ideas and techniques in such fields as psychology, the various branches of biology, medicine, the social sciences, industrial research and operations, and marketing was a striking phenomenon.

2. During the same period, persons interested in these other developments had founded a number of new societies, among them the Institute of Mathematical Statistics, the Econometric Society, the Psychometric Society and the American Society for Quality Control. All these societies were strongly concerned with statistical techniques, but none of them had any formal relation to the ASA.
3. The membership of the ASA was increasing and might be expected to grow rapidly in the post-war years. In 1945 there were about 3,300 members, at present there are close to 5,000.
4. With the formation of the United Nations, some of its agencies might be expected to foster new developments in international statistics.

In considering the future of the Association in the light of these factors, two principal choices appeared to be open. The Association might continue to give primary attention to applications in economics, leaving applications in other fields to be taken care of by other societies. This would have been a reasonable course of action. Although the Association had received an influx of members whose interests were in other fields, the primary concern of over half the members in 1945 was still with applications to economics or business, as revealed by the 1945 *Directory*.

The second course, the one actually adopted, was to try to give the Association a *central* role with regard to all fields of application of statistics. This decision was advocated by almost all members whose opinions were sought. It was a wise decision from many points of view, particularly when no one knew where important statistical applications might turn up next, when statistical activities were being parcelled out amongst numerous societies and when a strong national body might accomplish much in cooperation with international agencies. We should recognize, however, that the decision involved a real sacrifice, at least for a time, by the members in economics and business, since a relatively homogeneous society catering satisfactorily to them

was to be changed into something more amorphous whose future course was harder to predict. These members accepted and encouraged the change with excellent spirit and with, as might be expected, occasional grumbles.

SOME PROVISIONS OF THE 1948 CONSTITUTION

The decision having been taken, the new Constitution was constructed so as to introduce a number of devices that would make the desired changes easier to accomplish. I would like to describe the purposes, as I understand them, of some of the principal provisions in the 1948 Constitution.

Associated and affiliated societies. One of the most difficult questions was: what was to be the relation between the ASA and the other societies dealing with some aspect of statistics that had come into being or might be established in the future? Much thought was given to this question, including a study of various mechanisms that had been adopted by other large central organizations. Finally, it was decided to try two provisions, called *association* and *affiliation.*

Any other society interested in the objects of the ASA may apply to become an Associated or an Affiliated Society. The status of an Associated Society is intended for societies whose interest in statistics is strong: that of an Affiliated Society was intended to cover a looser type of connection, but since this provision was dropped in our recent minor revision of the Constitution, I will not go into detail about it. Proposals for association are examined by our Board of Directors and Council before a decision is taken to grant the status.

Each Associated Society receives the right to appoint two members to the Council of the ASA, one member to the editorial board of *The American Statistician*, and one member to the ASA Committee on Publications for each periodical which it publishes. The ASA is required to offer its publications to the members of Associated Societies on the same basis as to ASA members, and vice versa.

The arrangement involves a slight loss of autonomy by the ASA. In return, it establishes a definite method of liaison, makes our Council more representative of statistical interests as a whole, and puts us in a better position to play the kind of central role that was considered desirable.

Sections and section committees. If the ASA is to be a society whose members have a great variety of interests, what can be done to ensure that each of the principal interest-groups within the membership participates to its own satisfaction?

For dealing with this problem, the ASA had a successful precedent

in the Biometrics Section, which had been in existence for a number of years. Although only a small fraction of the membership was interested in biometry as such, this Section arranged programs at each annual meeting, held joint sessions at the meetings of a number of the biological societies and published the *Biometrics Bulletin* with financial backing from the ASA.

The 1948 Constitution encouraged the formation of Sections in other broad areas by providing for the establishment of *Section Committees.* The general function of Section Committees is "to further the development of statistics in fields not adequately covered at present by associated or affiliated societies." (Article X, 8). These Committees are represented on the ASA program committee in order to arrange programs in their individual areas. In course of time, a Section Committee may draw up a charter which on approval leads to the formation of a Section. The new Constitution looks still further ahead by providing that when a Section has grown large enough, the Section Committee may take the initiative in organizing an Associated Society.

Districts and District Committees. In nation-wide societies that are small, meetings tend to be on a national level. As the society grows in numbers, it becomes feasible to hold regional meetings which give more of the members a chance to participate. In the ASA we have been fortunate in having a long tradition of meetings both at the national level and through our Chapters at the local level. In order to encourage activities and meetings at an intermediate regional level, the Constitution provides for the setting-up of geographical districts. In each, there is a District Committee, with two members from each ASA chapter and from each local unit, if there are any, of any Associated or Affiliated Society. The District Committees thus provide a means for coordinating the activities of the ASA and related societies at both the local and regional levels.

Council. Finally, in order to give the membership a broader representation in the administration of the ASA, the Constitution created a new policy-making body, the Council. This consists of the Board of Directors, the editor of each ASA publication, two representatives from each district and one from each Section Committee with more than 75 members, as well as representatives of Associated Societies and an equal number of representatives-at-large. The Board of Directors, which in former times was the governing body, now serves as the executive committee of the Council. During 1953, the Council had 34 members, as compared with 13 on the Board.

THE ASSOCIATION'S EXPERIENCE UNDER THE 1948 CONSTITUTION

I would now like to describe how the new devices have operated during the past 5 years. In cases where things have not as yet worked quite as actively as was hoped, I do not want to give the impression of washing dirty linen in public, which would be most reprehensible for a President. My defense would be that this linen is not dirty, and it is not being washed, but merely aired.

Associated Societies. Up to the present time, only one organization has become linked to us through this provision—the East North American Region of the Biometric Society, which might be regarded as one of our own children grown up, since the Biometric Society is a natural outgrowth of our Biometrics Section.

This modest beginning is not surprising, because no strenuous efforts have been made to bring the provision to the attention of other societies. In my opinion, it is advisable to wait until the ASA has settled down under the new Constitution before exploring with some of the other societies the possibility of a closer relationship, although we have progressed far enough so that any good opportunity for initiating discussions should not be missed. Perhaps the most propitious times will be when cooperation has already arisen about some matter of mutual interest, or when a new society has been launched with the guidance of the ASA. With the older societies, we may also have to recognize and handle tactfully a problem of prestige. Some members of these societies may feel that Association implies in some way a recognition of a lower status. No such status was intended in framing these provisions, under which the ASA sacrifices some autonomy, but the other society does not, as is clearly stated in our Constitution.

Sections and Section Committees. Excellent progress has been made in establishing a well-rounded group of Sections. This year, the Section on Social Statistics has been added to those on Biometrics, Business and Economic Statistics and Training in Statistics. A Committee on Statistics in the Physical Sciences has been at work for 2 years. Jointly these 5 areas appear comprehensive enough to cover the major interests of practically all our members, at least for the time being. Perhaps the largest single group unrepresented by a Section are the members whose primary interest is in statistical theory. So long as the Institute of Mathematical Statistics continues to meet with us, as it has done consistently in the past, such members are unlikely to regard themselves as neglected. In arranging the large number of sessions (currently around 50) which now comprise the program at the annual

meeting, the Section representatives have worked most efficiently and amicably, and I believe that we have a smooth mechanism for accomplishing this complicated task. The Section Committees have also been active in varying degrees in other projects, and have been called upon on numerous occasions for advice by the Board and Council.

Districts and District Committees. Activity in arranging meetings of something approaching a regional character, which was one of the primary intentions in setting up districts, has proceeded satisfactorily. The initiative, however, has come from different directions on different occasions. The interesting programs at the United Nations headquarters in New York in 1952 and 1953 were a joint venture by several Chapters. The successful series of Institutes at the Universities of of Illinois and Pennsylvania and at the Carnegie Institute of Technology involved cooperative planning among a number of groups, prominent among them being the Business and Economic Statistics Section. The regional meeting to be held in San Francisco in December, 1954, will be the responsibility of the Western District. Thus, what was perhaps the principal object in setting up District Committees is being achieved, although the Committees themselves have not been uniformly active.

The Council. In creating the Council, the intent was to give the membership a larger role in the policy-making of the ASA and perhaps also to allow for more deliberation on policy problems. I think it is fair to say that these aims have not been fulfilled thus far. The annual meeting of the Council takes place at the beginning of the new President's term of office, a day or two after the new Council members have been elected. The agenda is a full one, with enough questions calling for immediate decision to leave little time or energy for leisurely discussion of long-range policy problems. The Board members tend to be the more active participants in the discussion, because they are more familiar with the issues than those who are not Board members.

It can be argued, of course, that if affairs are running smoothly without intense Council activity, as they appear to be, there is no point in looking for more work for the Council just to keep them busy. Also, a group with around 30 members is of an awkward size for some types of work and deliberation. The Council can meet at other times and can be polled by mail, so that it stands ready when any important policy matter arises. On the other hand, since the council is our policy-making body, our most representative body, and the body on which nominees from other societies will see us in action, there is a strong case for trying

to make it more continuously effective. There are several techniques that would be worth experimentation, and the Board has been considering a plan of action. I am sorry that during my term of office I did not make a beginning.

THE PRESENT STRUCTURE OF THE ASSOCIATION

As indicated previously, the wording of the 1948 Constitution suggests that the ASA would assume a more definitely central role in statistics by establishing, through association, links with other societies which recognized this role for the ASA. Section Committees were apparently regarded as more of an *interim* mechanism, since the Constitution describes them as applicable to "fields not adequately covered at present by associated or affiliated societies" and regards them as a means for organizing an associated society.

As events have turned out, the formation of Sections and Section Committees has been the predominant feature in the development of the ASA during the past five years, while only a bare beginning has been made in linking ourselves with other societies. This has been a sound order of procedure, in that we have been working hard to try to serve the whole range of statistics, before putting forward claims that we are able to do so. It now looks as if many of our most important activities during the next few years will be in the hands of the Sections. I hope that members of Section committees will realize how important these committees have become. Their useful activity is by no means confined to helping with the program at the Annual Meetings, but may include the planning of more specialized meetings, contributions to the publication program of the ASA and factual studies of problems that confront the content fields.

As the Sections become larger and better established, what will be the next step in the evolution of the ASA? In particular, what will happen if a Section develops into Associated Society ~~or if~~ a society already in existence in the field of the Section becomes associated with us? I do not know the answer, but some recent experiences of the Biometrics Section are worth noting.

After the North American regions of the Biometric Society had been established, the members of the Biometrics Section began a lively discussion of the future of this Section. Some members contended that the Biometrics Section should be dissolved. They claimed that the new regions of the Biometric Society could take care of the welfare of biometry in this country, that their administration would to a large

extent be in the hands of ASA members anyway, and that continuation of the Biometrics Section would be an unnecessary duplication of effort.

An opposing view was that for a statistician, membership in the Biometric Society serves a different purpose from membership in the Biometrics Section. At present, about half the members of the Biometric Society are biologists. If this Society is to flourish in its original objectives, it must continue to attract to membership a large number, preferably a majority, of biologists who would not join any statistical association. Thus the Biometric Society gives the statistician the opportunity to talk with *biologists*, learning their problems, working with them, and presenting new techniques for criticism and use. The ASA, on the other hand, is the place where statisticians in biometry can talk with *statisticians* in other content fields, both to find out what new techniques have developed in these fields and to present new ideas in biometry. From this point of view there was a strong argument for continuing the Biometrics Section as a nucleus for attracting future biometricians into the ASA, for cooperating with other Sections and for organizing programs on new or recent discoveries, where the technical level would be too high for most biologists.

After much debate, the decision was taken to continue the Biometrics Section. I do not claim that it was the argument given above which carried the day. Biometricians, like other statisticians, are fond of nice logical distinctions, and each tends to put forward a slightly different reason for advocating the same decision, and to attach great importance to the superiority of his reason over anyone else's, even though to an outsider the reasons are practically indistinguishable. But I hope that the argument will not be overlooked if other Sections blossom into full societies and their members are uncertain whether to continue the Section. If this concept of the purpose of a Section is sound, the greatest benefit will be obtained from the present ASA structure only if there is sustained cooperation among Sections and if members make a habit of attending sessions of several different Sections. There is, of course, nothing to prevent a member from belonging to every Section.

If the opposing view prevails, and if we are to look forward to seeing the Sections disband one by one as Associated Societies are formed (as might happen if there is a general lack of interest in continuing the Sections) then the structure of the ASA will evolve towards something different. A conservative might comment that it would then resemble either a jellyfish or an octopus, depending on how one looks at it. More

seriously, I do not mean to suggest that Sections should be kept alive if there is no intrinsic life in them. We should, however, have to re-examine the whole problem of the best type of structure for the ASA under the changed conditions. Actually, some types of organization that did not involve Sections at all were examined in the initial work for the 1948 Constitution, but were rejected as being unsuitable in our present state of growth.

SOME QUESTIONS CONCERNING THE VITALITY OF THE ASSOCIATION

To consider our present structure from a slightly different point of view, I would now like to pose a few broad questions which bear upon what might be called the state of health of the Association.

Can the ASA maintain the enthusiastic support of its members? Any large and heterogeneous society is likely to find that it is nobody's darling, because the affections of the members are accorded to some smaller and more homogeneous group in which they feel more at home. As the Association grows larger in its new role, it may be more difficult to give the members a real sense of participation. The *Journal* and *The American Statistician*, as the most tangible benefits from membership, have an important part to play, and it is currently planned to supplement these periodicals from time to time with special monographs and other publications of interest to the members. Meetings of a local or regional character are a beneficial addition to our Annual Meetings as a means of bringing together more of our members. Our Chapters and Sections may accomplish much in giving members a more immediate focus for their interests. Continued joint activity by different Sections will avoid a partitioning into self-contained groups that has occurred in some societies. In addition, I hope that members will continue to agree that statistics needs an all-embracing society, and will appreciate that the Association will inevitably become more diffuse as it succeeds in adopting this role.

Can the ASA continue to recruit young members? It is relatively painless for them to enter into membership: students pay only half the regular dues, as do also members under 30 during their first year. The office conducts a continuing campaign to spread information about membership, the groups approached being varied from year to year. As in other societies, our office finds that nothing succeeds so well as a personal approach from a present member, so that it is to our members and to the quality of our publications that we must look mainly for a steady recruitment of young persons.

Does the structure of the ASA encourage younger members, as they ma-

ture, to participate in the running of the ASA? Since the rapid growth of statistics is recent, we suffer relatively little from government by the grey-haired. Nevertheless, many of our most experienced members are heavily burdened with activities on behalf of scientific societies. For this reason, as well as to keep us supplied with fresh points of view, the talents of younger members should be utilized to the fullest extent. The Chapters and the Section and District Committees provide the first opportunity for younger members to undertake responsible tasks. For service at the national level on the Council or Board, the problem of introducing new blood is more difficult. In the elections, which are by majority vote, my impression is that the candidate who is more widely known (and usually older) is very frequently the winner. Something can be done about this problem both by the Committee on Elections when they nominate candidates and by the President when he appoints committees.

Is the ASA able to stimulate new developments in statistics? Some members have expressed the opinion that in the thirties and early forties the ASA missed an opportunity by not playing a more prominent part in the developments which led to the formation of a number of other societies with statistical interests. I am not sure that I would agree. In the Biometric Society, which we did help to establish, I have been slightly disturbed in case the statisticians should play too prominent a role relative to the biologists. In founding this kind of a society, there is something to be said for leaving much of the initiative to the scientists in the subject-matter field, who would not in general be members of the ASA. Nevertheless, our assumption of the role of a central organization with very wide interests does carry more responsibility for helping such developments, rather than leaving them to take place outside the ASA.

Here again we must rely mainly on the Section Committees, particularly when they arrange programs, to be on the lookout for new developments. Inspection of the wide range of our programs in recent years suggests that the committees have been lively and enterprising in this respect. The Board and Council and the office can also help. For a time, the Board felt impelled to adopt a cautious policy owing to our financial difficulties, but fortunately these appear to be well out of the way.

Is the ASA able to exercise leadership for statistics as a whole? So far as the use of statistics in government is concerned, our leadership is recognized as a result of a long history of disinterested service to agencies of

the government. I believe that international statistical agencies would also join in this recognition.

How do we stand in other areas involving statistical interests? Are we active enough in exercising leadership? These questions are more troublesome. Two areas that have always been of deep concern to the Council and Board are that involving relations with the public and that involving Statistical Standards. A piece of sound and important statistical work may be subjected to unjustified public attack, or a piece of shoddy and unscrupulous work, masking as statistically sound, may threaten to bring discredit on the profession. Should such circumstances arise, I imagine that most members would expect the Association to take corrective action. The problem of doing this effectively raises numerous difficulties. The critical moment for taking action may not be clear: there may be varying opinions about the most appropriate type of action; and the pressure of time may prevent thorough study of the issue before something must be done. For these reasons I am doubtful whether reliance on any standby body, such as the Council or some designated committee, will be adequate. The analogy with a fire brigade is not good, because nobody rings the alarm bell to tell us when to spring into action. The Council and Board have been struggling to consider what program of study might be initiated in order to establish a set of principles and a mode of action for dealing with such emergencies so that we will not be caught unawares. This is a task that needs all the help that members can give. For many of the problems it seems clear that to be fully effective, the ASA must work along with other societies that have statistical interests. Consequently, a program of this kind may be one means of drawing us closer to these societies.

Finally, any account of our present structure must recognize that we are a voluntary organization. Apart from a tiny office staff, everything that we do depends on the voluntary labor of the members. The Association can become what the members want it to be: there is no entrenched bureaucracy to impose its own pattern. Any member with a bright idea will receive an interested hearing (although he may sometimes have to talk a little loudly in order to do so). If his idea is bright enough, he will very likely find himself asked to carry it out as an enterprise of the Association. Secondly, we are a scientific as distinct from a professional society, in the sense that the Association has always worked for the highest statistical standards rather than for the economic interests of its members. As we grow larger, it may be harder to

retain this voluntary, scientific character while representing effectively the whole range of statistical activities. For my part, I hope that we can do both.

To summarize, the ASA is in a difficult period of growth in trying to keep up with an extraordinary expansion of statistics which scarcely anyone could have predicted accurately. In particular, the increasing specialization within statistics has set up forces which tend to decrease the amount of common interest amongst members and to split them into separate groups. The task of serving all areas of application in this rapidly-changing environment will require us to be wide-awake, adaptable, and receptive to new ideas and new ventures. My own appraisal would be that during the past five years our Association has made gratifying progress, especially in view of the financial stringencies which inflation imposed upon us. Some of the provisions of the 1948 Constitution have had only modest effects as yet, but these provisions have not proved harmful: they create mechanisms that will increase our flexibility in adapting ourselves to the future growth of the field of statistics. Although much remains to be done, I believe that we now have an organizational pattern that at least for the near future will enable us to take full advantage of our broad, common interests while giving scope also to our more specialized interests.

PRINCIPLES OF SAMPLING*

William G. Cochran, *Johns Hopkins University*
Frederick Mosteller, *Harvard University,*
John W. Tukey, *Princeton University*

I. SAMPLES AND THEIR ANALYSES

1. *Introduction*

Whether by biologists, sociologists, engineers, or chemists, sampling is all too often taken far too lightly. In the early years of the present century it was not uncommon to measure the claws and carapaces of 1000 crabs, or to count the number of veins in each of 1000 leaves, and then to attach to the results the "probable error" which would have been appropriate had the 1000 crabs or the 1000 leaves been drawn at random from the population of interest. Such actions were unwarranted shotgun marriages between the quantitatively unsophisticated idea of sample as "what you get by grabbing a handful" and the mathematical precise notion of a "simple random sample." In the years between we have learned caution by bitter experience. We insist on some semblance of mechanical (dice, coins, random number tables, etc.) randomization before we treat a sample from an existent population as if it were random. We realize that if someone just "grabs a handful," the individuals in the handful almost always resemble one another (on the average) more than do the members of a simple random sample. Even if the "grabs" are randomly spread around so that every individual has an equal chance of entering the sample, there are difficulties. Since the individuals of grab samples resemble one another *more* than do individuals of random samples, it follows (by a simple mathematical argument) that the means of grab samples resemble one another *less* than the means of random samples of the same size. From a grab sample, therefore, we tend to *under*estimate the variability in the population, although we should have to *over*-estimate it in order to obtain valid estimates of variability of grab sample means by substituting such an estimate into the formula for the variability of means of simple random samples. Thus using simple random sample formulas for grab sample means introduces a double

* This paper will constitute Appendix G of Cochran, Mosteller, and Tukey, *Statistical Problems of the Kinsey Report*, to be published by the American Statistical Association later this year as a monograph. The main body of this monograph was published in the *Journal* last December (Vol. 48 (1953), pp. 673–716).

bias, both parts of which lead to an unwarranted appearance of higher stability.

Returning to the crabs, we may suppose that the crabs in which we are interested are all the individuals of a wide-ranging species, spread along a few hundred miles of coast. It is obviously impractical to seek to take a simple random sample from the species—no one knows how to give each crab in the species an equal chance of being drawn into the sample (to say nothing of trying to make these chances independent). But this does not bar us from honestly assessing the likely range of fluctuation of the result. Much effort has been applied in recent years, particularly in sampling human populations, to the development of sampling plans which *simultaneously*,

(i) are economically feasible
(ii) give reasonably precise results, and
(iii) show within themselves an honest measure of fluctuation of their results.

Any excuse for the dangerous practice of treating non-random samples as random ones is now entirely tenuous. Wider knowledge of the principles involved is needed if scientific investigations involving samples (and what such investigation does not?) are to be solidly based. Additional knowledge of techniques is not so vitally important, though it can lead to substantial economic gains.

A botanist who gathered 10 oak leaves from each of 100 oak trees might feel that he had a fine sample of 1000, and that, if 500 were infected with a certain species of parasites, he had shown that the percentage infection was close to 50%. If he had studied the binomial distribution he might calculate a standard error according to the usual formula for random samples, $p \pm \sqrt{pq/n}$, which in this case yields $50 \pm 1.6\%$ (since $p = q = .5$ and $n = 1000$). In this doing he would neglect three things:

(i) Probable selectivity in selecting trees (favoring large trees, perhaps?),
(ii) Probable selectivity in choosing leaves from a selected tree (favoring well-colored or, alternatively, visibly infected leaves perhaps), and
(iii) the necessary allowance, in the formula used to compute the standard error, for the fact that he has not selected his leaves individually at random, as the mathematical model for a simple random sample prescribes.

Most scientists are keenly aware of the analogs of (i) and (ii) in their own fields of work, at least as soon as they are pointed out to them.

Far fewer seem to realize that, even if the trees were selected at random from the forest and the leaves were chosen at random from each selected tree, (iii) must still be considered. But if, as might indeed be the case, each tree were either wholly infected or wholly free of infection, then the 1000 leaves tell us *no more* than 100 leaves, one from each tree. (Each group of 10 leaves will be all infected or all free of infection.) In this case we should take $n = 100$ and find an infection rate of $50 \pm 5\%$.

Such an extreme case of increased fluctuation due to sampling in clusters would be detected by almost all scientists, and is not a serious danger. But less extreme cases easily escape detection and may therefore be very dangerous. This is one example of the reasons why the principles of sampling need wider understanding.

We have just described an example of *cluster sampling*, where the individuals or sampling units are not drawn into the sample independently, but are drawn in clusters, and have tried to make it clear that "individually at random" formulas do not apply. It was not our intention to oppose, by this example, the use of cluster sampling, which is often desirable, but only to speak for proper analysis of its results.

2. *Self-weighting probability samples*

There are many ways to draw samples such that each individual or sampling unit in the population has an equal chance of appearing in the sample. Given such a sample, and desiring to estimate the population average of some characteristic, the appropriate procedure is to calculate the (unweighted) mean of all the individual values of that characteristic in the sample. Because weights are equal and require no obvious action, such a sample is *self-weighting*. Because the relative chances of *different* individuals entering the sample are known and compensated for (are, in this case, equal), it is a *probability sample*. (In fact, it would be enough if we knew somewhat less, as is explained in Section 5.)

Such a sample need not be a simple random sample, such as one would obtain by numbering all the individuals in the population, and then using a table of random numbers to select the sample on the basis: one random number, one individual. We illustrate this by giving various examples, some practical and others impractical.

Consider the sample of oak leaves; it might in principle be drawn in the following way. First we list all trees in the forest of interest, recording for each tree its location and the number of leaves it bears. Then we draw a sample of 100 trees, arranging that the probability

of a tree's being selected is proportional to the number of leaves which it bears. Then on each selected tree we choose 10 leaves at random. It is easy to verify that each leaf in the forest has an equal chance of being selected. (This is a kind of two-stage sampling with probability proportional to size at the first stage.)

We must emphasize that such terms as "select at random," "choose at random," and the like, always mean that some mechanical device, such as coins, cards, dice, or tables of random numbers, is used.

A more practical way to sample the oak leaves might be to list only the locations of the trees (in some parts of the country this could be done from a single aerial photograph), and then to draw 100 trees in such a way that each tree has an equal chance of being selected. The number of leaves on each tree is now counted and the sample of 1000 is prorated over the 100 trees in proportion to their numbers of leaves. It is again easy to verify that each leaf has an equal chance of appearing in the sample. (This is a kind of two-stage sampling with probability proportional to size at the second stage.)

If the forest is large, and each tree has many leaves, either of these procedures would probably be impractical. A more practical method might involve a four-stage process in which:

(a) the forest is divided into small tracts,
(b) each tract is divided into trees,
(c) each tree is divided into recognizable parts, perhaps limbs, and
(d) each part is divided into leaves.

In drawing a sample, we would begin by drawing a number of tracts, then a number of trees in each tract, then a part or number of parts from each tree, then a number of leaves from each part. This can be done in many ways so that each leaf has an equal chance of appearing in the sample.

A different sort of self-weighting probability sample arises when we draw a sample of names from the Manhattan telephone directory, taking, say, every 17,387th name in alphabetic order starting with one of the first 17,387 names selected at random with equal probability. It is again easy to verify that every name in the book has an equal chance of appearing in the sample (this is a systematic sample with a random start, sometimes referred to as a systematic random sample).

As a final example of this sort, we may consider a national sample of 480 people divided among the 48 states. We cannot divide the 480 cases among the individual states in proportion to population very well, since Nevada would then receive about one-half of a case. If we

group the small states into blocks, however, we can arrange for each state or block of states to be large enough so that on a pro rata basis it will have at least 10 cases. Then we can draw samples within each state or block of states in various ways. It is easy to verify that the chances of any two persons entering such a sample (assuming adequate randomness within each state or block of states) are approximately the same, where the approximation arises solely because a whole number of cases has to be assigned to each state or block of states. (This is a rudimentary sort of stratified sample.)

All of these examples were (at least approximately) self-weighting probability samples, and all yield honest estimates of population characteristics. *Each one* requires a *different* formula for assessing the stability of its results! Even if the population characteristic studied is a fraction, almost never will

$$p \pm \sqrt{\frac{pq}{n}}$$

be a proper expression for "estimate ± standard error." In every case, a proper formula will require more information from the sample than merely the overall percentage. (Thus, for instance, in the first oak leaf example, the variability from tree to tree of the number infested out of 10 would be needed.)

3. *Representativeness*

Another principle which ought not to need recalling is this: By sampling we can learn only about collective properties of populations, not about properties of individuals. We can study the average height, the percentage who wear hats, or the variability in weight of college juniors, or of University of Indiana juniors, or of the juniors belonging to a certain fraternity or club at a certain institution. The population we study may be small or large, but there must be a population—and what we are studying must be a population characteristic. By sampling, we cannot study individuals as particular entities with unique idiosyncrasies; we can study regularities (including typical variabilities as well as typical levels) in a population as exemplified by the individuals in the sample.

Let us return to the self-weighted national sample of 480. Notice that about half of the times that such a sample is drawn, there will be no one in it from Nevada, while almost never will there be anyone from Esmeralda County in that state. Local pride might argue that "this

proves that the sample was unrepresentative," but the correct position seems to be this:

(i) the particular persons in the sample are there by accident, and this is appropriate, so far as population characteristics are concerned,

(ii) the sampling plan is representative since each person in the U.S. had *an equal chance* of entering the sample, whether he came from Esmeralda County or Manhattan.

That which can be and should be representative is the *sampling plan*, which includes the manner in which the sample was drawn (essentially a specification of what other samples might have been drawn and what the relative chances of selection were for any two possible samples) *and* how it is to be analyzed.

However great their local pride, the citizens of Esmeralda County, Nevada, are entitled to representation in a national sampling plan only as individual members of the U.S. population. They are *not* entitled to representation as a group, or as particular individuals—only as individual *members* of the U.S. population. The same is true of the citizens of Nevada, who are represented in only half of the actual samples. The citizens of Nevada, as a group, are no more and no less entitled to representation than *any* other group of equal size in the U.S. whether geographical, racial, marital, criminal, selected at random, or selected from those not in a particular national sample.

It is clear that many such groups fail to be represented in any particular sample, yet this is not a criticism of that sample. Representation is not, and should not be, by groups. It is, and should be, by individuals as *members* of the sampled population. Representation is not, and should not be, in any particular sample. It is, and should be, in the sampling *plan*.

4. *One method of assessing stability*

Because representativeness is inherent in the sampling plan and not in the particular sample at hand, we can never make adequate use of sample results without some measure of how well the results of this particular sample are likely to agree with the results of other samples which the same sampling plan might have provided. The ability to assess stability fairly is as important as the ability to represent the population fairly. Modern sampling plans concentrate on both.

Such assessment must basically be in terms of sample results, since these are usually our most reliable source of information about the population. There is no reason, however, why assessment should de-

pend only on the sample size and the overall (weighted) sample mean for the characteristic considered. These two suffice when measuring percentages with a simple random sample, but in almost all other cases the situation is more complex.

It would be too bad if, every time such samples were used, the user had to consult a complicated table of alternative formulas, one for each plan, before calculating his standard errors. (These formulas do need to be considered whenever we are trying to do a really good job of maximum stability for minimum cost—considered very carefully in selecting one complex design in preference to another.) Fortunately, however, this complication can often be circumvented.

One of the simplest ways is to build up the sample from a number of independent subsamples, each of which is self-sufficient, though small, and to tabulate the results of interest separately for each subsample. Then variation among separate results gives a simple and honest yardstick for the variability of the result or results obtained by throwing all the samples together. Such a sampling plan involves *interpenetrating replicate subsamples.*

All of us can visualize interpenetrating replicate subsamples when the individuals or sampling units are drawn individually at random. Some examples in more complex cases may be helpful. In the first oak leaf example, we might select randomly, not one sample of 100 trees, but 10 subsamples of 10 trees each. If we then pick 10 leaves at random from each tree, placing them in 10 bags, one for each subsample, and tabulate the results separately, bag by bag, we will have 10 interpenetrating replicate subsamples. Similarly, if we were to pick 10 subsamples out of the Manhattan phone book, with each subsample consisting of every 173,870th name (in alphabetic order) and with the 10 lead names of the 10 subsamples selected at random from the first 173,870 names we would again have 10 interpenetrating replicate subsamples.

We can always analyze 10 results from 10 independent interpenetrating replicate subsamples just as if they were 10 random selected individual measurements and proceed similarly with other numbers of replicate subsamples.

5. *General probability samples*

The types of sample described in the last section are not the only kinds from which we can confidently make inferences from the sample to the population of interest. Besides the trivial cases where the sample amounts to 90% or even 95% of the population, there is a broad class

of cases, including those of the last section as special cases. This is the class of *probability samples*, where:

(1) There is a population, the *sampled population*, from which the sample is drawn, and each element of which has some chance of entering the sample.

(2) For each pair of individuals or sampling units which are in the actual sample, the relative chances of their entering the sample are known. (This implies that the sample was selected by a process involving one or more steps of mechanical randomization.)

(3) In the analysis of the actual sample, these relative chances have been compensated for by using relative weights such that

(relative chance) *times* (relative weight) equals a constant.

(4) For any two possible samples, the sum of the reciprocals of the relative weights of all the individuals in the sample is the same.

(Conditions (3) and (4) can be generalized still further.) In practice of course, we ask only that these four conditions shall hold with a sufficiently high degree of approximation.

We have made the sampling plan representative, not by giving each individual an equal chance to enter the sample and then weighting them equally, but by a more noticeable process of compensation, where those individuals very likely to enter the sample are weighted less, while those unlikely to enter are weighted more when they do appear. The net result is to give each individual an equal chance of affecting the (weighted) sample mean.

Such general probability samples are just as honest and legitimate as the self-weighting probability samples. They often offer substantial advantages in terms of higher stability for lower cost.

We can alter our previous examples, so as to make them examples of general, and not of self-weighting, probability samples. Take first the oak leaf example. We might proceed as follows:

(1) locate all the trees in the forest of interest,

(2) select a sample of trees at random,

(3) for each sampled tree, choose 10 leaves at random and count (or estimate) the total number of leaves,

(4) form the weighted mean by summing the products

(fraction of 10 leaves infested) *times*

(number of leaves on the tree)

and then divide by the total number of leaves on the 100 trees in the sample.

When we selected trees at random, each tree had an equal probability of selection. When we chose 10 leaves from a tree at random, the chance of getting a particular leaf was

$$\frac{10}{\text{(number of leaves on the tree)}}.$$

Thus the chance of selecting any one leaf was a constant multiple of this and was proportional to the *reciprocal* of the number of leaves of the tree. Hence the correct relative weight is proportional to the number of leaves on the tree, and it is simplest to take it as 1/10 of that number. After all, summing the products

(fraction of 10 infected) *times* (leaves on tree)

or

(1/10) *times* (number out of 10 infected) *times* (leaves on tree)

over all trees in the sample gives the same answer. One-tenth of this answer is given by summing

(1/10) *times* (number out of 1 infected) *times* (leaves on tree)

or

$$\text{(number out of 1 infected)}\ \frac{\text{(leaves on tree)}}{10}$$

which shows that the weighted mean prescribed above is just what would have been obtained with relative weights of (number of leaves on tree)/10.

If in sampling the names in the Manhattan telephone directory, we desired to sample initial letters from P through Z more heavily, we might proceed as follows:

(1) Select one of the first 17,387 names at random with equal probability as the lead name.
(2) Take the lead name, and every 17,387th name in alphabetic order following it, into the sample.
(3) Take every name which begins with $P, Q, R, S, \cdots, Z$ *and* is the 103rd or 207th name after a name selected in step 2 of the sample.

Each name beginning with $A, B, \cdots, N, O$ has a chance of 1/17,387 of entering the sample. Each name beginning with $P, Q, \cdots, Y, Z$ has a chance of 3/17,387 of entering the sample (it enters if any one of *three* names among the first 17,387 is selected as the lead name). Thus the relative weight in the sample of a name beginning with $A, B, \cdots, N, O$ is 3 times that of a name beginning with $P, Q, \cdots, Y, Z$. The weighted mean is found simply as:

$$\frac{3(\text{sum for } A, B, \cdots, N, O\text{'s}) + (\text{sum for } P, Q, \cdots, Y, Z\text{'s})}{3(A, B, \cdots, N, O\text{'s in sample}) + (P, Q, \cdots, Y, Z\text{'s in sample})}.$$

Finally we may wish to distribute our national sample of 480 with 10 in each state. The analysis exactly parallels the oak leaf case, and we have to form the sum of

(mean for state sample) *times* (population of state)

and then to divide by the population of the U.S.

6. *Nature and properties of general probability samples*

We can carry over the use of independent interpenetrating replicates to the general case without difficulty. We need only remember that the replicates must be independent. In the oak leaf example, the replicates must come from groups of independently selected trees. In the Manhattan telephone book example, the replicates must be based on independently chosen lead names; in the national sample, the replicates must have members in every state. In every case they must interpenetrate, and do this independently.

It is clear from discussion and examples that general probability samples are inferior to self-weighting probability samples in two ways, for both simplicity of exposition and ease of analysis are decreased! If it were not for compensating advantages, general probability samples would not be used. The main advantages are:

(1) better quality for less cost due to reduction in administrative costs or prelisting cost,

(2) better quality for less cost because of better allocation of effort over strata,

(3) greater possibility of making estimates for individual strata.

All three of these advantages can be illustrated on our examples. In the general oak leaf example, in contrast to the first oak leaf example in Section 2, there is no need to determine the size (number of leaves) of all trees. This is a clear cost reduction, whether in money or time. Suppose that, in the Manhattan telephone book sample, one aim was an opinion study restricted to those of Polish descent. Such persons' names tend to be concentrated in the second part of the alphabet, so that the general sample will bring out more persons of Polish descent and the interviewing effort will be better allocated. In the case of the national sample of 480, the general sample, although probably giving a less stable national result, does permit (rather poor) state-by-state estimates where the self-weighting sample would skip Nevada about half the time.

It is perhaps worth mentioning at this point that, if cost is proportional to the total number of individuals without regard to number of strata

or the distribution of interviews among strata, the optimum allocation of interviews is proportional to the product

(size of stratum) *times* (standard deviation within stratum).

In particular, optimum allocation calls for sample strata not in proportion to population strata. If we weight appropriately, disproportionate samples will be better than proportionate ones—if we choose the disproportions wisely.

In specifying the characteristics of a probability sampling at the beginning of this paper, we required that there be a *sampled population*, a population from which the sample comes and each member of which has a chance of entering the sample. We have not said whether or not this is exactly the same population as the population in which we are interested, the *target population*. In practice they are rarely the same, though the difference is frequently small. In human sampling, for example, some persons cannot be found and others refuse to answer. The issues involved in this difference between sampled population and target population are discussed at some length in Part II, and in chapter III-D of Appendix D in our complete report.

7. *Stratification and adjustment*

In many cases general probability samples can be thought of in terms of

(1) a subdivision of the population into strata,
(2) a self-weighting probability sample in each stratum, and
(3) combination of the stratum sample means weighted by the size of the stratum.

The general Manhattan telephone book sample can be so regarded. There are two strata, one made up of names beginning in $A, B, \cdots, N, O$, and the other made up of names beginning in $P, Q, \cdots, Y, Z$. Similarly the general national sample may be thought of as made up of 48 strata, one for each state.

This manner of looking at general probability samples is neat, often helpful, and makes the entire legitimacy of unequal weighting clear in many cases. But it is not general. For in the general oak leaf example, if there were any strata they would be whole trees or parts of trees. And not all trees were sampled. (Still every leaf was fairly represented by its equal chance of affecting the weighted sample mean.) We cannot treat this case as one of simple stratification.

The stratified picture is helpful, but not basic. It must fail as soon as there are more potential strata than sample elements, or as soon as

the number of elements entering the sample from a certain stratum is not a constant of the sampling plan. It usually fails sooner. There is no substitute for the relative chances that different individuals or sampling units have of entering the sample. This is the basic thing to consider.

There is another relation of stratification to probability sampling. When sizes of strata are known, there is a possibility of *adjustment*. Consider taking a simple random sample of 100 adults in a tribe where exactly 50% of the adults were known to be males and 50% females. Suppose the sample had 60 males and 40 females. If we followed the pure probability sampling philosophy so far expounded, we should take the equally weighted sample mean as our estimate of the population average. Yet if 59 of the 60 men had herded sheep at some time in their lives, and none of the 40 women, we should be unwise in estimating that 59% of the tribe had herded sheep at some time in their lives. The adjusted mean

$$.50\left(\frac{59}{60}\right) + .50\left(\frac{0}{40}\right) = 49+\%$$

is a far better indicator of what we have learned.

How can adjustment fail? Under some conditions the variability of the adjusted mean is enough greater than that of the unadjusted mean to offset the decrease in bias. It may be a hard choice between adjustment and nonadjustment.

The last example was extreme, and the unwise choice would be made by few. But, again, less extreme cases exist, and the unwise choice, whether it be to adjust or not to adjust, may be made rather easily (and probably has been made many times). A quantitative rule is needed. One is given in chapter V-C of the complete report. In the preceding example the relative sizes of the strata were known exactly. It turns out that inexact knowledge can be included in the computation without great increase in complexity.

An example in Kinsey's area is cited by one critic of the Kinsey report:

> These weighted estimates do not, of course, reflect any population changes since 1940, which introduces some error into the statistics for the present total population. Moreover, on some of the very factors that Kinsey demonstrates to be correlated with sexual behavior, there are no Census data available. For example, religious membership is shown to be a factor affecting sexual behavior, but Census data are lacking and no weights are assigned. While the investigators interviewed members of various religious groups, there is no assurance that each group is proportionately represented, because of the lack of systematic sampling controls. Thus, the proportion of

Jews in Kinsey's sample would seem to be at least 13 per cent whereas their true proportion in the population is of the order of 4 per cent.[1]

Do we know the percentage of Jews well enough to make an adjustment for it? If we can assess the stability of the "4%" figure, the procedure of Chapter V-C will answer this question. Failing this technique, we could translate the question into more direct terms as follows: "In considering Kinsey's results, do we want to have 13 per cent Jews or 4 per cent Jews in the sampled population?" and try to answer with the aid of general knowledge and intuition.

We have discussed the adjustment of a simple random sample. The same considerations apply to the possibility of adjusting any self-weighting or general probability sample. No new complications arise when adjustment is superposed on weighting. The presence of a complication might be suspected in the case where not all segments appear in the sample, and we attempt to use these segments as strata. Careful analysis shows the absence of the complication, as may be illustrated by carrying our example further.

Suppose that the sheep-herding tribe in question contains a known, very small percentage of adults of indeterminate sex, and that none have appeared in our sample. To be sure, their existence affected, albeit slightly, the chances of males and females entering the sample, but it does not affect the thinking which urged us to take the adjusted mean. We still want to adjust, and have only the question "Adjust for what?" to answer.

If the fraction of indeterminate sex is 0.000002, and the remainder are half males and half females, and if our anthropological expert feels that about 1 in 7 of the indeterminate ones has herded sheep, we have a choice between

$$.499999\left(\frac{59}{60}\right) + .499999\left(\frac{0}{40}\right) + .000002\left(\frac{1}{7}\right)$$

which represents adjustment for three strata, one measured subjectively, and

$$.500000\left(\frac{59}{60}\right) + .500000\left(\frac{0}{40}\right)$$

which represents adjustment for the two observed strata.

Clearly, in this extreme example, the choice is immaterial. Clearly,

[1] Hyman, H. H. and Sheatsley, P. B. "The Kinsey report and survey methodology," *International Journal of Opinion and Attitude Research*, Vol. 2 (1948), 184–85.

also, the estimated accuracy of the anthropologist's judgment must enter. We can again use the methods of Chapter V-C.

8. *Upper semiprobability sampling*

Let us be a little more realistic about our botanist and his sample of oak leaves. He might have an aerial photograph, and be willing to select 100 trees at random. But any ladder he takes into the field is likely to be short, and he may not be willing to trust himself in the very top of the tree with lineman's climbing irons. So the sample of 10 leaves that he chooses from each selected tree, will not be chosen at random. The lower leaves on the tree are more likely to be chosen than the highest ones.

In the two-stage process of sampling, the first stage has been a probability sample, but the second has not (and may even be entirely unplanned!). These are the characteristic features of an *upper semiprobability sample.* As a consequence, the sampled population agrees with the target population in certain large-scale characteristics, but not in small-scale ones and, usually, not in other large-scale characteristics. Thus, if in the oak leaf example we use the weights appropriate to different sizes of tree, as we should, the sampled population of leaves will

(1) have the correct relative number of leaves for each tree, but

(2) will have far too many lower leaves and far too few upper leaves.

The large-scale characteristic of being on a particular tree is a matter of agreement between sampled and target populations. The large-scale characteristic of height in the tree (and many small-scale characteristics that the reader can easily set up for himself), is a matter of serious disagreement between sampled and target populations.

The sampled population differs from the target population within each segment, here a tree, although sampled population segments and target population segments are in exact proportion.

If infestation varies between the bottoms and the tops of the trees, this type of sampling will be biased, and, while the inferences from sample to sampled population will be correct, they may be useless or misleading because of the great difference between sampled population and target population.

Such dangers always exist with any kind of nonprobability sampling. Upper semiprobability sampling is no exception. By selecting the trees at random we have stultified biases due to probable selectivity between trees, and this is good. But we have done nothing about almost certain selectivity between leaves on a particular tree—this may be all right, or very bad. It would be nice to always have probability

samples, and avoid these difficulties. But this may be impractical. (The conditions under which a nonprobability sample may reasonably be taken are discussed in Part II.)

There is one point which needs to be stressed. The change from probability sampling within segments (in the example, within trees) to some other type of sampling, perhaps even unplanned sampling, shifts a large and sometimes difficult part of the inference from sample to target population—shifts it by moving the sampled population away from the target population toward the sample—shifts it from the shoulders of the statistician to the shoulders of the subject matter "expert." Those who use upper semiprobability samples, or other nonprobability samples, take a heavier load on themselves thereby.

Upper semiprobability samples may be either self-weighting or general. The "quota samples" of the opinion pollers, where interviewers are supposed to meet certain quotas by age, sex, and socioeconomic status, are rather crude forms of upper semiprobability samples, and are often self-weighting. Bias within segments arises, some contribution being due, for example, to the different availability of different 42 year old women of the middle class. The sampled population may contain sexes, ages, and socioeconomic classes in the right ratios, but retiring persons are under-represented (and hermits are almost entirely absent) in comparison with the target population.

Election samples of opinion, although following the same quota pattern, will ordinarily only be self-weighting within states (if we ignore the "who will vote" problem). Predictions are desired for individual states. If Nevada had a mere 100 cases in a self-weighting sample, the total size of a national sample would have to be about 100,000. When national percentages are to be compiled, it would be foolish not to weight each state mean in accordance with the size of the state. No one would favor, we believe, weighting each state equally just because there may be (and probably are) biases within each state.

Disproportionate samples and unequal weights are just as natural and wise a part of upper semiprobability sampling as they are of probability sampling. The difficulties of upper semiprobability sampling do not lie here; instead they lie in the secret and insidious biases due to selectivity within segments.

Our sampling of names from the Manhattan telephone directory might conceivably be drawn by listing the numbers called by subscribers on a certain exchange during a certain time, and then taking into the sample names from each exchange in proportion to the names listed for the exchange. The result would be an upper semiprobability sample

with substantial selectivity within the segments, which here are exchanges. The nature of this selectivity would depend on the time of day at which the listing was made.

Whether all segments are represented in an upper semiprobability sample or not, the segments may be used as strata for adjustment. The situation is exactly similar to that for probability sampling. The only difficulty worthy of note is the difficulty of assessing the stability of the various segment means.

Independent interpenetrating replicate subsamples can be used to estimate stabilities of over-all or segment means in upper semiprobability samples without difficulty, if we can obtain a reasonable facsimile of independence in taking the different subsamples. They provide, if really independent, respectable bases for inference from sample to sampled population. We still have a nonprobability sample, however, and there is no reason for the sampled population to agree with the target population. The problem is just reduced to "What was the sampled population?"

What finally is the situation with regard to bias in an upper semiprobability sample? We shall have a weighted mean or an adjusted one. In either case, any bias originally contributed by selectivity between segments will have been substantially removed. *But*, in either case, the contribution to bias due to selectivity within segments will *remain unchanged*. This is an unknown and hence additionally dangerous, sort of bias.

The great danger in weighting or adjusting such samples is not so much that that weighting or adjusting may make the results worse (as it will from time to time) but rather that its use may cause the user to feel that his values are excellent because they are "weighted" or "adjusted" and hence to neglect possible or likely biases within segments. Like all other nonprobability sample results, weighted means from upper semiprobability samples should be *presented and interpreted with caution.*

9. *Salvage of unplanned samples*

What can we do for such samples? We can either try to improve the results of their analysis, or try to inquire how good they are anyway. We may try to improve either actual quality, or our belief in that quality. The first has to be by way of manner of weighting or adjustment, the second must involve checking sample characteristics against population characteristics.

Weighting is impossible, since we cannot construct a sampling plan and hence cannot estimate chances of entering the sample in any other manner than by observing the sample itself. So all that we can do under this head is to adjust. We recall the salient points about adjustment, which are the same in a complete salvage operation as they are in any other situation:

(1) The population is divided into segments.
(2) Each individual in the sample can be uniquely assigned to a segment.
(3) The population fraction is either known with inappreciable error or estimated with known stability.
(4) The procedures of Chapter V-C of Appendix C of the complete report are applied to determine whether, or how much, to adjust.

After adjustment, what is the situation as to bias? Even worse than with upper semiprobability sampling, because if we do not adjust, we cannot escape bias by turning to weighting. In summary

(1) whether adjusted or not, the result contains all the effects of all the selectivity exercised *within* segments, while
(2) if adjustment is refused by the methods of Chapter V-C, we face additional biases resulting from selectivity *between* segments of a magnitude comparable with the difference between unadjusted and adjusted mean.

This is, to put it mildly, not a good situation.

Clearly even more caution is needed in presenting and interpreting the results of a salvage operation on an unplanned sample than for any of the other types of sample discussed previously. (If it were not for the psychological danger that adjustment might be regarded as cure, the caution required for results based on the original, unadjusted, unplanned sample would, however, be considerably greater.)

Having adjusted or not as seems best, what else can we do? Only something to make ourselves feel better about the sample. Some other characteristic than that under study can sometimes be compared in the adjusted sample and in the population. A large difference is evidence of substantial bias within segments. Good agreement is comforting, and strengthens the believability of the adjusted mean for the characteristic of interest. The amount of this strengthening depends very much on the *a priori* relation between the two characteristics.

Some would say that an unplanned sample does not deserve adjustment, but the discussion in Part II indicates that if any sort of a summary is to be made, it might as well, in principle, be an adjusted mean.

II. SYSTEMATIC ERRORS

In order to understand how systematic errors in sampling should be treated, it seems both necessary and desirable to fall back on the analogy with the treatment of systematic errors in measurement. No clear account of the situation for sampling seems to be available in the literature, although understanding of the issues is a prerequisite to the critical assessment of nonprobability samples. On the other hand, one of physical science's greatest and more recurrent problems is the treatment of systematic errors.

10. *The presence of systematic errors*

Almost any sort of inquiry that is general and not particular involves both sampling and measurement, whether its aim is to measure the heat conductivity of copper, the uranium content of a hill, the visual acuity of high school boys, the social significance of television or the sexual behavior of the (white) human (U.S.) male. Further, *both* the measurement *and* the sampling will be imperfect in almost every case. We can define away either imperfection in certain cases. But the resulting appearance of perfection is usually only an illusion.

We can define the thermal conductivity of a metal as the average value of the measurements made with a particular sort of apparatus, calibrated and operated in a specified way. If the average is properly specified, then there is no "systematic" error of measurement. Yet even the most operational of physicists would give up this definition when presented with a new type of apparatus, which standard physical theory demonstrated to be less susceptible to error.

We can relate the result of a sampling operation to "the result that would have been obtained if the same persons had applied the same methods to the whole population." But we want to know about the population and not about what we would find by certain methods. In almost all cases, applying the method to the "whole" population would miss certain persons and units.

Recognizing the inevitability of (systematic) error in both measurement and sampling, what are we to do? Clearly, attempt to hold the combined effect of the systematic errors down to a reasonable value. What is reasonable? This must depend on the cost of further reduction and the value of accurate results. How do we *know* that our systematic errors have been reduced sufficiently? We don't! (And neither does the physicist!) We use all the subject-matter knowledge, information and semi-information that we have—we combine it with whatever internal evidence of consistency it seems worthwhile to arrange

for the observations to provide. The result is not foolproof. We may learn new things and do better later, but who expects the last words on any subject?

In 1965, a physicist measuring the thermal conductivity of copper would have faced, unknowingly, a very small systematic error due to the heating of his equipment and sample by the absorption of cosmic rays, then unknown to physics. In early 1946, an opinion poller, studying Japanese opinion as to who won the war, would have faced a very small systematic error due to the neglect of the 17 Japanese holdouts, who were discovered later north of Saipan. These cases are entirely parallel. Social, biological and physical scientists all need to remember that they have the same problems, the main difference being the decimal place in which they appear.

If we admit the presence of systematic errors in essentially every case, what then distinguishes good inquiry from bad? Some reasonable criteria would seem to be:

(1) Reduction of exposure to systematic errors from *either* measurement *or* sampling to a level of unimportance, *if possible and economically feasible*, otherwise

(1+) Balancing the assignment of available resources to reduction in systematic or variable errors in either measurement or sampling reasonably well, in order to obtain a reasonable amount of information for the "money."

(2) Careful consideration of possible sources of error and careful examination of the numerical results.

(3) Presentation of results and inferences in a manner which adequately points out both observed variability and conjectured exposure to systematic error.

In many situations it is easy, and relatively inexpensive, to reduce the systematic errors in sampling to practical unimportance. This is done by using a probability sampling plan, where the chance that any individual or other primary unit shall enter the sample is known, and allowed for, and where adequate randomness is ensured by some scheme of (mechanical) randomization. The systematic errors of such a sample are minimal, and frequently consist of such items as:

(a) failure of individuals or primary units to appear on the "list" from which selection has been made,

(b) persons perennially "not at home" or samples "lost,"

(c) refusals to answer or breakdowns in the measuring device.

These are the hard core of causes of systematic error in sampling. Fortunately, in many situations their effect is small—there a prob-

ability sample will remove almost all the systematic error due to sampling.

11. *Should a probability sample be taken?*

But this does not mean that it is always good policy to take probability samples. The inquirer may not be able to "afford" the cost in time or money for a probability sample. The opinion pollers do not usually afford a probability sample (instead of designating individuals to be interviewed by a random, mechanical process, they allow their interviewers to select respondents to fill "quotas") and many have criticized them for this. Yet the behavior of the few probability samples in the 1948 election (see pp. 110–112 of *The Pre-election Polls of 1948*, Social Science Research Council Report No. 60) does not make it clear that the opinion pollers should spend their limited resources on probability samples for best results. (Shifts *toward* a probability sample have been promised, and seem likely to be wise.)

The statement "he didn't use a probability sample" is thus *not* a criticism which should end further discussion and doom the inquiry to the cellar. It is always necessary to ask two questions:

(a) Could the inquirer afford a probability sample?

(b) Is the exposure to systematic error from a non-probability sample small enough to be borne?

If the answer is "no" to both, then the inquiry should not be, or have been, made—just as would be the case with a physical inquiry if the systematic errors of all the forms of measurement which the physicist could afford were unbearably large.

If the answer is "yes" to the first question and "no" to the second, then the failure to use a probability sample is very serious, indeed.

If the answer is "yes" to both, then careful consideration of the economic balance is required—however it should be incumbent on the inquirer using a nonprobability sample to show why it gave more information per dollar or per year. (As statisticians, we feel that the onus is on the user of the *non*probability sample. Offhand we know of no expert group who would wish to lift it from his shoulders.)

If the answer is "no" to the first question, and "yes" to the second, then the appropriate reaction would seem to be "lucky man."

Having admitted that the sampling, as well as the measurement, will have some systematic errors, how then do we do our best to make good inferences about the subject of inquiry? Sampling and measurement being on the same footing, we have only to copy, for the sampling area, the procedure which is well established and relatively well understood for measurement. This procedure runs about as follows:

We admit the existence of systematic error—of a difference between the quantity measured (the measured quantity) and the quantity of interest (the target quantity). We ask the observations about the measured quantity. We ask our subject matter knowledge, intuition, and general information about the relation between the measured quantity and the target quantity.

We can repeat this nearly verbatim for sampling:

We admit the existence of systematic error—of a difference between the population sampled (the sampled population) and the population of interest (the target population). We ask the observations about the sampled population. We ask our subject matter knowledge, intuition, and general information about the relation between sampled population and target population.

Notice that the measured quantity is not the raw readings, which usually define a *different* measured quantity, but rather the adjusted values resulting from all the standard corrections appropriate to the method of measurement. (Not the actual gas volume, but the gas volume at standard conditions!) Similarly, the result for the sampled population is not the raw mean of the observations, which usually defines a *different* sampled population, but rather the adjusted or weighted mean, all corrections, weightings and the like appropriate to the method of sampling having been applied. Weighting a sample appropriately is no more fudging the data than is correcting a gas volume for barometric pressure.

The third great virtue of probability sampling is the relative definiteness of the sampled population. It is usually possible to point the finger at most of the groups in the target population who have no chance to enter the sample, who therefore were not in the sampled population; and to point the finger at many of the groups whose chance of entering the sample was less than or more than the chance allotted to them in the computation, who therefore were fractionally or multiply represented in the sampled population. When a nonprobability sample is adjusted and weighted to the best of an expert's ability, on the other hand, it may still be very difficult to say what the sampled population really is. (Selectivity *within* segments cannot be allowed for by weights or adjustments, but it arises to some extent in every nonprobability sample and alters the sampled population.)

12. *The value and conditions of adjustment*

Some would say that correcting, adjusting and weighting most nonprobability samples is a waste of time, since you do not *know*, when this process has been completed, to what sampled population the

adjusted result refers. This is entirely equivalent to saying that it does not pay to adjust the result of a physical measurement for a known systematic error because there are, undoubtedly, other systematic errors and some of them are likely to be in the other direction. Let us inquire into good practice in the measurement situation, and see what guidance it gives us for the sampling situation.

When will the physicist adjust the principle for the known systematic error? When (i) he has the necessary information and (ii) the adjustment is likely to help. The necessary information includes a theory or empirical formula, and the necessary observations. Empirical formulas and observations are subject to fluctuations, so that adjustment will usually change the magnitude of fluctuations as well as altering the systematic error. The adjustment is likely to help unless the supposed reduction of systematic error coincides with a substantial increase in fluctuations.

If the known systematic error is so small as not to

(1) affect the result by a meaningful amount, or

(2) affect the result by an amount likely to be as large as, or a substantial fraction of, the unknown systematic errors,

then the physicist will report either the adjusted or the unadjusted value. If he reports the unadjusted value, he should state that the adjustment has been examined, and is less than such-and-so. To do this, either he must have calculated the adjustment or he must have had generally applicable and strong evidence that it is small.

In any event, his main care, which he will not always take, must be to warn the reader about the dangers of further systematic errors, perhaps, in some cases, even by saying bluntly that "the adjusted value isn't much better than the raw value," and then provide raw values for those who wish to adjust their own.

If the physicist is aware of systematic errors of serious magnitude and has no basis for adjustment, his practice is to name the measured quantity something, like Brinnell hardness, Charpy impact strength, or if he is a chemist—iodine value, heavy metals as Pb, etc. By analogy, those who feel that the combination of recall and interview technique make Kinsey's results subject to great systematic error might well define "KPM sexual behavior" as a standard term,[2] and work with this.

By analogy then, when should a nonprobability sample be adjusted in principle? (Most probability samples are made to be weighted anyway—this is part of the design and must be carried out.) When (i)

[2] The letters KPM stand for Kinsey, Pomeroy and Martin, the authors of *Sexual Behavior in the Human Male.*

we have the necessary information and (ii) when the adjustment is likely to help. The necessary information will usually consist of facts or estimates of the true fractions in the population of the various segments.

When is the adjustment likely to help? This problem has usually been a ticklish point requiring technical knowledge and intuition. A quantitative solution is now given in Chapter V–C of Appendix C in the complete report. With this as a guide, it should be possible to make reasonable decisions about the helpfulness of adjustment.

If the decision is to adjust, we should accept the sampled population corresponding to the adjusted mean, and calculate the adjustment. We then report the adjusted value unless the adjustment is small, when we may report the unadjusted value with the statement that the adjustment alters it by less than such-and-so.

Our main care, which we may not always take, must be to warn the reader about the dangers of further lack of representativeness, perhaps, in some cases, even by saying bluntly that "the adjusted mean isn't much better than the raw mean, even if we took 20 pages to tell you how we did it and six months to do it," and to provide raw means for those who wish to adjust their own.

If we were prepared to report an unadjusted mean, we were clearly inviting inference to some sampled population. Adjustment will give us a sampled population that is usually nearer to the target population. Hence we should adjust.

If we cannot adjust, and must present raw data which we feel *badly* needs adjustment, we may say that this is what we found in these cases—take 'em or leave 'em. Except from the point of view of protecting the reader from over-belief in the results, this would seem to be a counsel of despair. By analogy with the physicist, it seems better to introduce "KPM sexual behavior" and its analogs in such situations.

DO PERSONS LOST TO LONG TERM OBSERVATION HAVE THE SAME EXPERIENCE AS PERSONS OBSERVED?

EVALUATION OF ANTISYPHILITIC THERAPY*

THEODORE J. BAUER, JAMES F. DONOHUE, VINCENT LARSEN, ALBERT P. ISKRANT, AND QUENTIN R. REMEIN†

MEDICAL research has long been faced with the problem of patients lapsing from observation during the evaluation of treatment. To date methods of analysis of results of therapy have assumed that the patients who lapse from observation would have had the same experience as those who remained under observation. To the extent that this assumption is not correct the calculation of the results of therapy is in error. It has been claimed that the "observed" is weighted in favor of failures as relapses would return for more treatment while those who are getting along all right would not bother to return for posttreatment observation. On the other hand it has been stated that the failures go elsewhere for retreatment instead of to the source of the original treatment and thus bias a study in favor of "good" results.

In an effort to test the hypothesis underlying current evaluation of treatment (viz., the "not observed" would have had the same experience as the observed) a special study in the evaluation of therapy for syphilis, the Blue Star Research Study, was initiated by the Division of Venereal Disease, U. S. Public Health Service, in which an attempt was made to hold a group of patients to 100 per cent follow-up and compare the results of therapy with those obtained when no intensive follow-up effort was made [1]. These patients included 560 persons with secondary syphilis confirmed by darkfield examination who had had no previous antisyphilitic therapy of any kind. The follow-up effort was over 90 per cent effective over a period of two years, largely due to the work of the physicians in the cooperating treatment facilities and specially trained research investigators whose sole functions

* Presented before the session on Statistical Evaluation of Clinical Data. American Statistical Association Annual Meeting, December 28, 1951.

† Dr. Theodore J. Bauer formerly Chief, Division of Venereal Disease, U.S. Public Health Service, represents the physicians and nurses who, over the years, devoted their diligent efforts to the treatment and observation of Blue Star Research Study patients. Mr. Donohue, Principal Statistician, and Mr. Larsen, Health Program Representative, represent the research investigators whose tireless and careful work in selecting and holding patients to observation, have made possible the collection of the data utilized in this paper. Mr. Iskrant, formerly Principal Statistician of the Division of Venereal Disease, and Mr. Remein, Statistician, Division of Venereal Disease, prepared this statistical analysis.

have been to assist the physicians in selecting patients and to see that these patients are held to follow-up observation.

The determination of response to therapy in syphilis requires observation of the patient for a long period of time following treatment. In the primary or secondary stage, relapse following treatment generally occurs within two years, if at all, although it may take as long as twenty years or more to determine whether late, disabling effects occur (such as paresis, tabes dorsalis, and aneurysm). This study is limited to a two-year evaluation of response to therapy for secondary syphilis. The presence of relapse is determined by physical examination, darkfield microscopy of sera from lesions, and blood and spinal fluid serologic tests. Results of the blood test are usually reported quantitatively in titer units based on successive twofold dilutions of the patient's serum. When treatment is successful, the titer gradually declines until negativity is reached a number of months after treatment. After the initial schedule of treatment is completed, no further therapy is administered unless lesions reappear, unless a serologic relapse evidenced by a sustained rise in serologic titer occurs, or a high titer (32 Kahn units or more) is sustained for one year or more following therapy [1]. Death from syphilis rarely occurs in the first few years after onset.

The criteria (other than diagnostic) taken into consideration in the selection of patients for the Blue Star Research Study are willingness of the patient to cooperate in long term follow-up, good general health, residential stability, and, in a measure, personal stability (i.e., alcoholics, drug addicts, etc. were excluded if known). It was felt that these characteristics would not affect the relapse rate to an appreciable extent. In spite of screening, however, some such unstable persons were included in the study because their characteristics were not known at the time of selection.

In the following pages the experience of these intensively followed patients is used in two different methods to test the validity of the hypothesis that patients who lapse from observation would have had the same experience as those who remained under observation, first, by analysis of the results for patients within the special study and second, by comparison of results of special study patients with results for another group of patients who were followed less intensively.

METHOD 1

This analysis is limited to patients treated for secondary syphilis who were selected for the special intensive follow-up study. Three

schedules of treatment, all utilizing crystalline penicillin G are included.

1. A total of 2,800,000 units of aqueous penicillin alone—25,000 units administered every three hours for fourteen days.
2. A total of 2,800,000 units of aqueous penicillin with conjunctive arsenoxide and bismuth—25,000 units of penicillin administered every three hours for 14 days and a total of 4 to 6 mg. of arsenoxide per kilogram of body weight (or a total of 300 mg. in persons weighing over 60 kg.) and 600 mg. of bismuth.
3. A total of 3,400,000 units of aqueous penicillin—40,000 units administered every two hours for seven days.

Almost all of the patients treated on these schedules have had the opportunity to be observed for at least two years, and sufficient cases for evaluation have been accumulated. Patients were observed monthly for the first year and quarterly during the second year.

Follow-up of the patients was secured by the research investigator stationed at each cooperating facility. A variety of follow-up techniques was used by each investigator including letters, telegrams, telephone calls, visits to patients, etc. The cooperation given by public clinics and physicians in many parts of the country made follow-up continuity possible for many patients who moved beyond commuting distance from the treating facility. Not all patients proved to be cooperative. These cases would have been lost to follow-up had it not been for the diligence of the investigators. The success of follow-up is indicated in Table 1 where it can be seen that approximately 95 per

TABLE 1
SPECIAL STUDY IN TREATMENT OF SECONDARY SYPHILIS
POSTTREATMENT OBSERVATION SUCCESS, BY TREATMENT SCHEDULE

	2,800,000 units Penicillin Alone			2,800 000 units Pen., Ars. & Bis.			3,400 000 units Penicillin Alone			Total		
	Number		Per Cent	Number		Per Cent	Number		Per Cent	Number		Per Cent
Under Observation	160		95.8	179		92.7	147		95.5	486		94.6
Retreated before 2 years		30			29			16			75	
2 years or more since treatment		121			149			124			394	
Less than 2 years since treatment		9			1			7			17	
Lost to Observation	7		4.2	14		7.3	7		4.5	28		5.4
Lapsed		6			13			7			26	
Died (not assoc. with syphilis)		1			1			0			2	
Total number treated	167		100.0	193		100.0	154		100.0	514		100.0

cent of the patients under study were observed for two years or until retreatment, or are still under observation at this writing (about 4

years since the study began). No significant differences among the three schedules in the percentage of patients under observation were noted. (The probability of getting differences giving a chi-square greater than that obtained through chance alone is slightly more than 6 out of 10.)

In consultation with research investigators using detailed records on interviews with patients and on the amount of effort required to keep patients under observation, the patients have been classified according to whether or not they would probably have been lost to routine methods of follow-up. Patients who missed two or more appointments without a good reason; those who frequently required special attention in obtaining observations such as home visits, provision for transportation to the facility even when no observations were actually missed; and those who moved out of the area served by the facility—all these were considered as becoming lost to routine follow-up during the two year observation period. For each patient who would have been lost to routine follow-up, the posttreatment observation period in which he would have lapsed was estimated as closely as possible.

It was our original intention to divide the research cases into two groups, the cooperative and the uncooperative, and compare the cumulative retreatment rates in both groups. Unfortunately there was no way of determining which patients would be cooperative and which uncooperative except by observing the patient's behavior. Obviously the longer the time over which the patient was observed the greater the opportunity for showing uncooperativeness. Therefore, those patients who were retreated in the early posttreatment months did not have a chance to become uncooperative and hence in the early months the presumably cooperative patients showed higher retreatment rates.

We therefore decided to make our comparison on a more realistic basis by comparing the results of all patients in the study with the results that would have been obtained with the same cases if they had not had this concentrated follow-up. This would in essence amount to comparing the retreatment rate of patients without intensive follow-up to the retreatment rate of the same patients with intensive follow-up and would resemble the findings in actual practice.

The method of calculation of the retreatment rates is that used by the Division of Venereal Disease and previously described by several of us [3, 4]. Briefly, in this method the retreatment rate is calculated by making appropriate adjustment for the loss of patients from observation. The total used for computing retreatment, seropositivity, and seronegativity rates is adjusted by including the same proportion of

retreated as non-retreated patients remaining under observation. Then if a_n persons are observed in period n or later of whom c_n persons are retreated in period n, denoting the adjusted total cases by e,

$$e_n = \frac{a_n e_{(n-1)}}{a_{(n-1)} - c_{(n-1)}}.$$

The retreatment rate in period n is $100 \times c_n/e_n$, and the cumulative retreatment rate is the sum of the retreatment rates for all individual observation periods through n. The per cent seropositive or seronegative in period n is simply the number seropositive or seronegative divided by the adjusted total cases times 100.

The retreatment rate obtained by this method is the same as that obtained by the method commonly referred to as "the life table method." The seronegativity and seropositivity rates differ somewhat since "the life table method" cumulates sustained seronegativity rates whereas the method described here computes the seronegativity rate for each particular period based only on cases observed in that period or later.

The following two groups were tabulated and analyzed separately:

A—All patients in the study (i.e., intensive follow-up)—analysis of the results of therapy including all posttreatment observations through two years of follow-up on patients in the study.

B—Same patients with "routine" follow-up—analysis of the results of therapy for the same patients including only those observations on the uncooperative patients prior to the posttreatment period in which they would most likely have lapsed.

A couple of examples should help make clear precisely which observations were included in group A and which, in group B. Patient L.M. was treated for secondary syphilis and observed for six months. At the end of this time he moved to another city, 500 miles away from the treatment center. Arrangements were made for the patient to be observed by the local clinic in the new city of residence, and follow-up continued for the complete two-year period. In group A the results of all observations on this patient are included for the two years. In group B observations on this patient are included for only six months; that is, until he moved.

Or take another case, F. G., who was treated for secondary syphilis and was observed for 9 months following treatment. He did not appear for his tenth month examination. When the investigator found the patient and interviewed him, the patient indicated that he believed himself to be cured and had decided to stop coming in for examinations.

The patient did, however, agree to accompany the investigator to the clinic for an examination and serologic test. For the remainder of the two years, the investigator had to find the patient and bring him to the clinic for each examination. In group A all observations of this patient are included for the two years. In group B only the observations for nine months are included.

Tables 2, 3, and 4 and Figure 1 show by treatment schedule the comparative data on the results of therapy for the same patients for all observations and for those which would ordinarily be made without intensive follow-up. The greatest difference between the cumulative retreatment rates for both groups amounted to only 1.6 per cent at the seventh month on the 3,400,000 units of penicillin schedule. At no point are there significant differences[1] in rates between the two follow-up conditions (at the level of $P = .95$). At 21–24 months one schedule shows no difference in retreatment rates, and the others show differences in opposite directions. The largest difference was 1.1% in the 3,400,000 unit schedule. The probability of getting larger differences than this by chance alone is nearly one half.

---

[1] In testing for significance the intensively followed group has been considered the population consisting of N treated patients. In the ith observation period since treatment, s_i patients were retreated representing p_i proportion retreated. In the sample obtained through routine follow-up, primes are used to denote the estimates. To determine whether the cumulative proportion retreated over r observation periods in the sample differs significantly from the cumulative of the parameter over these same intervals, the following formulas are applied:

$$\sigma_{p_r} = \sqrt{\frac{\sum_{i=1}^{r} p_i\left(1 - \sum_{i=1}^{r} p_i\right)(N - E)}{E(N - 1)}}$$

where E, the effective sample size as adapted from Cornfield (2), is defined as the equivalent sample size in which there are no losses from observation but which yields the same proportion retreated and the same number of retreated cases as were observed; i.e.,

$$E = \frac{\sum_{i=1}^{r} s_i'}{\sum_{i=1}^{r} p_i'}.$$

The t-test is then applied as follows:

$$t = \frac{\sum_{i=1}^{r} p_i - \sum_{i=1}^{r} p_i'}{\sigma_{p_r}}.$$

TABLE 2

COMPARISON OF RESULTS OF THERAPY EVALUATION FOR INTENSIVE AND "ROUTINE" FOLLOW-UP

SECONDARY SYPHILIS—CUMULATIVE RETREATMENT RATES

TREATMENT SCHEDULE: Penicillin—2,800,000 units—25,000 every 3 hours "Crystalline G" (14 days) No arsenoxide, no bismuth

A—Intensive Follow-up

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
-1	—	—	—	166	99.4	1	0.6	167
1–2	2	1.2	1.2	144	86.2	21	12.6	167
2–3	—	—	1.2	118	71.1	46	27.7	166
3–4	3	1.8	3.0	87	52.4	74	44.6	166
4–5	1	0.6	3.6	64	38.6	96	57.8	166
5–6	5	3.0	6.6	50	30.1	105	63.3	166
6–7	3	1.8	8.4	40	24.2	111	67.3	165
7–8	1	0.6	9.0	34	20.6	116	70.3	165
8–9	2	1.2	10.2	30	18.2	118	71.6	165
9–10	2	1.2	11.4	24	14.8	120	73.8	163
10–11	3	1.8	13.2	21	12.9	120	73.8	163
11–12	1	0.6	13.8	18	11.1	122	75.0	163
12–15	3	1.8	15.6	13	8.0	124	76.2	163
15–18	2	1.2	16.8	8	5.0	126	78.0	162
18–21	2	1.3	18.1	9	5.7	119	76.0	157
21–24	—	—	18.1	8	5.4	113	76.3	148

B—Routine Follow-up

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
-1	—	—	—	164	99.4	1	0.6	165
1–2	2	1.3	1.3	135	85.4	21	13.3	158
2–3	—	—	1.3	109	70.3	44	28.4	155
3–4	3	2.0	3.3	79	53.1	65	43.7	149
4–5	1	0.7	4.0	59	40.2	82	55.9	147
5–6	5	3.5	7.5	43	30.1	89	62.4	143
6–7	2	1.4	8.9	35	24.9	93	66.2	140
7–8	1	0.7	9.6	29	20.8	97	69.6	139
8–9	2	1.4	11.0	26	18.8	97	70.1	138
9–10	2	1.5	12.5	20	14.8	98	72.6	135
10–11	2	1.5	14.0	17	13.0	95	72.9	130
11–12	—	—	14.0	15	11.6	96	74.3	129
12–15	2	1.7	15.7	10	8.3	91	75.9	120
15–18	2	1.7	17.4	6	5.2	90	77.4	116
18–21	1	0.9	18.3	7	6.2	85	75.4	113
21–24	—	—	18.3	6	5.6	81	76.0	107

* Adjusted total cases in each period includes the number of cases observed in this period or later and cases retreated in previous periods adjusted for losses from observation.

TABLE 3

COMPARISON OF RESULTS OF THERAPY EVALUATION FOR INTENSIVE AND "ROUTINE" FOLLOW-UP

SECONDARY SYPHILIS—CUMULATIVE RETREATMENT RATES

TREATMENT SCHEDULE: Penicillin—2,800,000 units—25,000 every 3 hours "Crystalline G" 4–6 mg. arsenoxide per kg. of body weight (or total of 300 mg. in persons weighing over 60 kg.) and 600 mg. of bismuth (14 days)

A—Intensive Follow-up

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
–1	—	—	—	192	99.5	1	0.5	193
1–2	—	—	—	178	92.7	14	7.3	192
2–3	1	0.5	0.5	140	73.7	49	25.8	190
3–4	2	1.1	1.6	108	56.8	79	41.6	190
4–5	2	1.1	2.7	79	41.8	105	55.6	189
5–6	3	1.6	4.3	61	32.3	120	63.5	189
6–7	4	2.1	6.4	48	25.5	128	68.1	188
7–8	—	—	6.4	42	22.3	134	71.3	188
8–9	2	1.1	7.5	36	19.2	138	73.4	188
9–10	1	0.5	8.0	35	18.8	136	73.2	186
10–11	1	0.5	8.5	30	16.2	139	75.3	185
11–12	3	1.6	10.1	26	14.1	140	75.8	185
12–15	4	2.2	12.3	20	10.9	141	76.8	184
15–18	2	1.1	13.4	15	8.3	141	78.3	180
18–21	3	1.7	15.1	9	5.0	143	79.9	179
21–24	1	0.6	15.7	6	3.4	142	80.9	175

B—Routine Follow-up

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
–1	—	—	—	187	99.5	1	0.5	188
1–2	—	—	—	161	91.5	15	8.5	176
2–3	1	0.6	0.6	122	72.6	45	26.8	168
3–4	2	1.3	1.9	85	53.8	70	44.3	158
4–5	2	1.4	3.3	57	39.4	83	57.4	145
5–6	2	1.4	4.7	41	29.2	93	66.2	141
6–7	3	2.2	6.9	31	22.9	95	70.2	135
7–8	—	—	6.9	29	22.1	93	71.0	131
8–9	1	0.8	7.7	21	16.9	94	75.5	125
9–10	—	—	7.7	21	17.0	93	75.8	123
10–11	1	0.8	8.5	18	14.7	94	76.8	122
11–12	3	2.5	11.0	14	11.9	91	77.1	118
12–15	2	1.8	12.8	10	9.0	87	78.2	111
15–18	2	1.9	14.7	6	5.7	83	79.5	104
18–21	—	—	14.7	4	4.0	81	81.2	100
21–24	1	1.0	15.7	1	1.0	82	83.2	99

* Adjusted total cases in each period includes the number of [illegible] observed in this period or later and cases retreated in previous periods adjusted for losses from observation.

TABLE 4

COMPARISON OF RESULTS OF THERAPY EVALUATION FOR INTENSIVE AND "ROUTINE" FOLLOW-UP

SECONDARY SYPHILIS—CUMULATIVE RETREATMENT RATES

TREATMENT SCHEDULE: Penicillin—3,400,000 units—40,000 every 2 hours "Crystalline G" (7 days), No arsenoxide, no bismuth

A—Intensive Follow-up

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
-1	—	—	—	153	99.4	1	0.6	154
1–2	1	0.7	0.7	145	94.8	7	4.6	153
2–3	—	—	0.7	122	79.7	30	19.6	153
3–4	3	2.0	2.7	98	64.1	51	33.3	153
4–5	3	2.0	4.7	79	51.6	67	43.8	153
5–6	—	—	4.7	71	46.4	75	49.0	153
6–7	4	2.6	7.3	59	38.6	83	54.2	153
7–8	1	0.7	8.0	51	33.3	90	58.8	153
8–9	2	1.3	9.3	39	25.7	99	65.2	152
9–10	—	—	9.3	32	21.1	106	69.8	152
10–11	—	—	9.3	29	19.2	108	71.6	151
11–12	—	—	9.3	24	15.9	113	74.9	151
12–15	1	0.7	10.0	19	12.8	115	77.4	149
15–18	—	—	10.0	19	12.8	115	77.4	149
18–21	1	0.7	10.7	14	9.5	118	80.0	148
21–24	—	—	10.7	9	6.5	115	83.0	139

B—Routine Follow-up

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
-1	—	—	—	147	99.3	1	0.7	148
1–2	1	0.7	0.7	133	95.0	6	4.3	140
2–3	—	—	0.7	103	78.7	27	20.6	131
3–4	2	1.6	2.3	78	63.5	42	34.2	123
4–5	2	1.7	4.0	61	51.4	53	44.6	119
5–6	—	—	4.0	55	47.1	57	48.8	117
6–7	2	1.7	5.7	47	41.0	61	53.2	115
7–8	1	0.9	6.6	39	35.0	65	58.3	111
8–9	2	1.9	8.5	29	27.6	67	63.8	105
9–10	—	—	8.5	24	23.1	71	68.3	104
10–11	—	—	8.5	20	19.9	72	71.6	101
11–12	—	—	8.5	14	14.1	77	77.4	100
12–15	—	—	8.5	12	12.6	75	78.8	95
15–18	—	—	8.5	12	12.8	74	78.7	94
18–21	1	1.1	9.6	8	8.6	76	81.7	93
21–24	—	—	9.6	5	5.6	75	84.7	89

* Adjusted total cases in each period includes the number of cases observed in this period or later and cases retreated in previous periods adjusted for losses from observation.

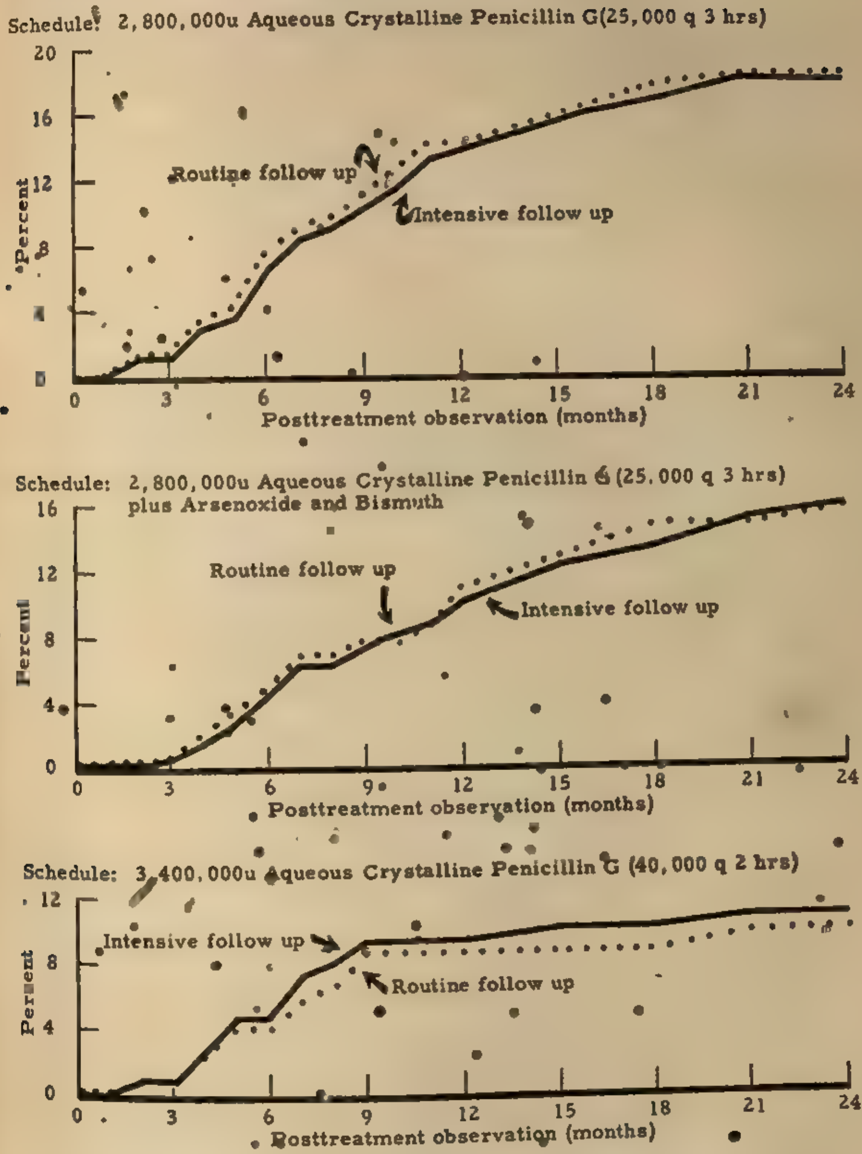


FIG. 1. Results of therapy evaluation—intensive and routine follow-up—secondary syphilis—cumulative retreatment rates.

If the intensively followed group is taken as the population and the routinely followed group as a sample, from that population, the remainder of the population consists of those patients who would have been lost to observation. Using the "effective-sample-size" notion [2], this situation closely approximates a condition where two samples are drawn from a population without replacement so that $n_1+n_2=N$. It can be readily demonstrated that, where $n_1+n_2=N$,

$$t=\frac{\bar{X}-\bar{X}_1'}{\sigma_{\bar{x}_1'}}=\frac{\bar{X}_2'-\bar{X}}{\sigma_{\bar{x}_2'}}=\frac{\bar{X}_2'-\bar{X}_1'}{\sigma_{\bar{x}_{1-2}'}},$$

that is, the t-value of the difference between two sample means where $n_1+n_2=N$ is the same as the t-value of the difference between either sample mean and the parameter. Applied to the present problem, this indicates that there are no significant differences between the results for persons remaining under observation and for those lost to observation since no significant differences were noted between the routinely followed and the intensively followed patients.

Table 5 shows the percentage of total cases followed for two years, the lowest limit being about 50 per cent followed for two years. Then with at least 50 per cent follow-up, patients lost to observation in the evaluation of treatment for syphilis, very likely have the same experience as patients remaining under observation.

TABLE 5

POSTTREATMENT OBSERVATION UNDER INTENSIVE AND ROUTINE CONDITIONS OF FOLLOW-UP

Special Study in Treatment of Secondary Syphilis

Treatment Schedule	A—Intensive Follow-up			B—Routine Follow-up		
	Total cases treated	Followed 21–24 mos. Number	Followed 21–24 mos. Per Cent	Total cases treated	Followed 21–24 mos. Number	Followed 21–24 mos. Per Cent
2,800,000 units Penicillin Alone	167	148	88.6	167	107	64.1
2,800,000 units Pen. with Ars. & Bis.	193	175	90.7	193	99	51.3
3,400,000 units Penicillin Alone	154	139	90.3	154	89	57.8

METHOD 2

In this method we have compared the results of treatment with penicillin in a group of patients with "routine" follow-up with the results among patients in the intensively followed group previously discussed. Note that in Method 1 we compared patients followed routinely with the *same* patients followed intensively; in Method 2 the

groups are mutually exclusive. The Division of Venereal Disease receives records from many hospitals and clinics on treatment and follow-up of patients treated for early syphilis. Many schedules of treatment are included in these records which are analyzed periodically for comparative evaluation. The results of these evaluations are published by the Division for program guidance. The follow-up on these patients is not as intensive as in the special study. All patients are advised of the importance of posttreatment follow-up before they are discharged from the treating facility. Form letters are sent to all patients reminding them when examinations are due, and occasionally, visits are made to the patients by a representative of the health department. In this routine method follow-up is conducted entirely on an impersonal basis, and there are many lapses from observation. In order to test the validity of a statistical evaluation with incomplete follow-up, a group of patients treated with crystalline penicillin G from the "routine" evaluation was compared with the patients treated with crystalline penicillin G in the intensive follow-up group. Both groups of patients were treated in the same clinics and during the same time periods (July 1946–December 1948).

Among patients with intensive follow-up the amount of penicillin ranged from 2,800,000 units to 4,200,000 units for an average of 3,130,000 units.[2] Among the patients with routine follow-up total dosage ranged from 2,400,000 units to 4,800,000 units for an average of 3,369,000 units.[2] Available evidence indicates that a difference of 239,000 units in an average penicillin dosage of over 3,000,000 units would make very little difference in the retreatment rates in the two groups inasmuch as the dosage-response curve has very little slope at 3,000,000 units and above. Therefore, the retreatment rates for the two groups can be expected to be approximately equal.

The intensive follow-up group includes 253 cases with 92.1 per cent observed for two years (or until retreated). The routine follow-up group had 1,864 patients treated of which 41.7 per cent were observed for two years (or until retreated). Results of therapy through 24 months are presented in Table 6. At the 24th month the cumulative retreatment rates are practically identical. Figure 2 presents a graphic comparison of the results. It can be concluded that the intensively followed cases were retreated earlier than those routinely followed, as evidenced by the slightly higher retreatment rate in the intensively followed group throughout the first year of observation. Also cases with intensive

[2] The geometric mean was used in this calculation because the dosage-response curve is such that arithmetic changes in the retreatment rate are associated with geometric changes in dosage.

follow-up are recorded as having reversed to negative more rapidly in the first six months than did cases routinely followed. Differences are negligible after the first six months of posttreatment observation. The fact that intensively followed cases were observed more frequently

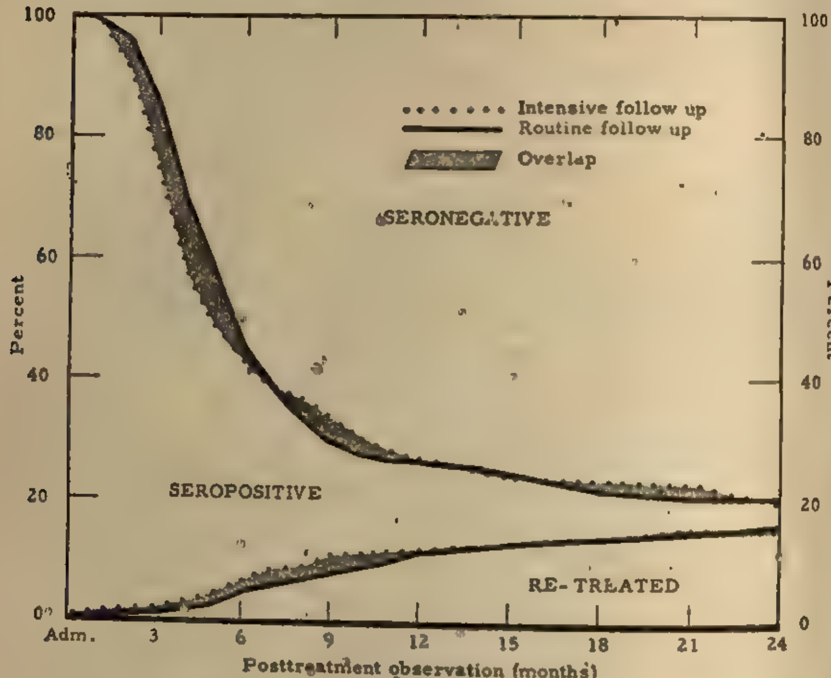


FIG. 2. Comparison of results of treatment with crystalline penicillin G of secondary syphilis in a group with intensive follow-up and in a group with routine follow-up.

than those under routine follow-up would account for the earlier detection of cases requiring retreatment among those intensively followed. This would probably also account for differences in the rapidity of reversal to seronegative. In spite of earlier retreatment and seronegativity rates, however the cumulative results from the 12th month on are almost identical for both groups.

CONCLUSION AND DISCUSSION

Two methods have been used to present evidence concerning the assumption that persons lost to observation have the same experience as persons observed in the evaluation of treatment for syphilis. In the

TABLE 6

RESULTS OF TREATMENT OF SECONDARY SYPHILIS WITH CRYSTALLINE PENICILLIN G IN A GROUP WITH INTENSIVE FOLLOW-UP AND IN A GROUP WITH "ROUTINE" FOLLOW-UP

Intensive Follow-up Group

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
-1	1	0.4	0.4	250	98.8	2	0.8	253
1-2	3	1.2	1.6	224	89.6	22	8.8	250
2-3	—	—	1.6	182	72.8	64	25.6	250
3-4	2	0.8	2.4	137	55.0	106	42.6	249
4-5	4	1.6	4.0	111	44.8	127	51.2	248
5-6	6	2.4	6.4	89	36.0	142	57.5	247
6-7	5	2.0	8.4	76	30.9	149	60.6	246
7-8	1	0.4	8.8	69	28.1	155	63.0	246
8-9	5	2.1	10.9	54	22.[illegible]	163	66.9	244
9-10	—	—	10.9	46	19.1	169	70.0	241
10-11	1	0.4	11.3	40	16.6	174	72.1	241
11-12	1	0.4	11.7	34	14.1	179	74.2	241
12-15	3	1.2	12.9	26	10.8	183	76.2	240
15-18	2	0.8	13.7	23	9.6	184	76.6	240
18-21	3	1.3	15.0	19	7.9	184	77.0	239
21-24	—	—	15.0	11	4.7	187	80.2	233

Routine Follow-up Group

Observation Period (months)	Retreated Cases			Not re-treated				Adjusted Total Cases*
				Seropositive		Seronegative		
	Number	Per Cent	Cumulative %	Number	Per Cent	Number	Per Cent	
-1	1	0.1	0.1	1,856	99.6	7	0.4	1,864
1-2	4	0.2	0.3	1,765	96.4	60	3.3	1,830
2-3	7	0.4	0.7	1,526	85.2	254	14.2	1,792
3-4	13	0.7	1.4	1,188	68.0	534	30.6	1,747
4-5	21	1.2	2.6	940	55.5	710	41.9	1,695
5-6	37	2.2	4.8	667	40.5	899	54.6	1,647
6-7	14	0.9	5.7	523	32.9	977	61.4	1,592
7-8	15	1.0	6.7	421	27.2	1,024	66.1	1,549
8-9	18	1.2	7.9	328	21.8	1,056	70.2	1,503
9-10	13	0.9	8.8	277	18.9	1,060	72.3	1,466
10-11	16	1.1	9.9	238	16.7	1,042	73.3	1,421
11-12	17	1.2	11.1	205	15.0	1,012	73.8	1,370
12-15	28	2.1	13.2	144	11.0	986	75.6	1,304
15-18	8	0.7	13.9	86	7.8	860	78.1	1,101
18-21	8	0.9	14.8	54	5.7	746	79.3	940
21-24	8	1.0	15.8	35	4.5	619	79.6	778

* Adjusted total cases in each period includes the number of cases observed in this period or later and cases retreated in previous periods adjusted for losses from observation.

first method, therapy results in a group of intensively followed patients were compared with the results which would have occurred among the same patients if intensive methods had not been used. In the second method therapy results in two mutually exclusive groups were compared. One group consisted of the previously mentioned patients who were followed intensively, and the other group consisted of routinely followed patients in the same treatment centers. No significant differences in retreatment rates were observed by either method, and it is our conclusion that in these series with at least 42 per cent complete follow-up at two years after treatment patients lost to observation had the same experience as those who remained under observation.

While these findings are of considerable value in the evaluation of therapy for syphilis, there is no evidence for their application to the study of therapy response in other diseases. For instance, where death from a disease frequently occurs after treatment, it is not at all likely that persons lost to observation would have the same experience as those observed.

REFERENCES

[1] Bauer, T. J., "Evaluation of antisyphilitic therapy with intensive follow-up, I. The plan," *Journal of Venereal Disease Information*, 32 (1951), 355–59.

[2] Cornfield, Jerome in Cutler, Sidney J., "Cancer illness among residents in Atlanta, Georgia, 1947," *National Cancer Institute of the National Institutes of Health, Cancer Morbidity Series*, No. 1, p. 30, 1950.

[3] Iskrant, A. P., Bowman, R. W., and Donohue, J. F., "Techniques in the evaluation of antisyphilitic therapy," *Public Health Reports*, 63 (1948), 965–77.

[4] Iskrant, A. P., Remein, Q. R., and Donohue, J. F., "Evaluation of antisyphilitic therapy with intensive follow-up, III. Statistical method of analysis and its critical evaluation," *Journal of Venereal Disease Information*, 32 (1951), 371–75.

APPLICATIONS OF STATISTICAL METHODS TO SEDIMENTARY ROCKS*

W. C. KRUMBEIN
Northwestern University

Statistical methods find wide application in geology, especially in the study of textures and composition of sedimentary rocks. Certain apparent irregularities in the data, such as highly skewed distributions, use of weight instead of number frequencies, use of unequal class intervals, and some others, required development of special methods of statistical analysis. In part, logarithmic transformations permitted application of conventional methods to the data. Some sedimentary attributes approach Gaussian distributions with no complicating factors. Mineral composition data are commonly binomial or Poisson distributions. Analysis of variance and experimental design are becoming increasingly important in further analysis of geological data.

INTRODUCTION

SEVERAL circumstances controlled the development of statistical thinking in geology. The first fields opened to statistical analysis, some 50 years ago, concerned data which did not seem to lend themselves to then current methods of analysis. Mineral composition of sediments, for example, yielded discrete rather than continuous distributions; and size frequency distributions of sediments, although continuous, required use of unequal class intervals and were commonly highly skewed. Moreover, a single sand sample may contain millions of grains, so that weight percentage instead of number of grains was more convenient for expressing frequency. Contemporary textbooks on statistics said little or nothing about handling such irregular data. As a result, techniques adapted to these special needs were developed directly by geologists.

Histograms of sand analyses were used before 1900, and since about 1925 logarithmic forms of histograms and cumulative curves have been in common use. The median and quartile deviation came into use about 1930, and the median is still the most popular average for reporting sedimentary grain size data. In part, the inertia of established procedure and the large amount of data published as median and quartile summaries have operated against use of more efficient statistics in ex-

* Paper presented before the Chicago meeting of the American Statistical Association, December 28, 1952. The manuscript has been partially revised to include developments during 1953.

pressing size analysis data. As long ago as 1935, Eisenhart [10] applied the Chi-square test to sampling problems in sedimentation, and numerous other workers, mentioned below, helped make available methods of wider and more general applicability than those initially introduced. Analysis of variance techniques have been used sporadically for about a decade, and in 1946 Swineford and Swineford [33] published a comprehensive study based on a three factor analysis of variance model.

Other fields of geology, notably paleontology and geomorphology, not faced initially with discrete or logarithmic distributions, adopted standard methods of analysis from the start. Applications in these fields have been expanded since the 1930's, and the process was accelerated by publication of Simpson and Roe's *Quantitative Zoology* in 1939 [32]. In 1948 and 1949 analysis of variance and multivariate analysis were applied in paleontology by Burma [3] and Miller [22]. In 1950 Strahler [34] studied relations between samples and populations of surface slopes in geomorphologic analysis.

Although modern statistical methods are being used to some extent in geology, the general state of statistical knowledge among geologists is rather unsatisfactory. Students seldom have courses in the subject, and some geology teachers perhaps tend to emphasize graphic procedures with little attention to underlying theory. It seems fair to state, however, that there is an expanding interest in the subject and a corresponding increase of appreciation of what statistics can and cannot do.

PROPERTIES OF SEDIMENTARY ROCKS

Inasmuch as this paper is addressed to an audience of statisticians, it seems appropriate to define the scope of the present subject. Sedimentary rocks are deposits of solid materials on the earth's surface produced by mechanical, chemical, or biological agencies in any medium (air, water, glacial ice) under normal conditions of the surface. All sedimentary rocks have attributes of composition, texture, and structure. Composition refers to mineralogical or chemical make-up of the rock; texture refers to characteristics of the grains or particles and the grain-to-grain relations among them; and structure refers to larger features of the deposit, such as stratification, geometrical attitude of the strata, and included organic remains.

Textural and compositional properties of sediments have been studied in more detail by statistical methods than have sedimentary structures, although many geological field studies involve the statistics of structures. Grain orientation, directions of cross-bedding, and atti-

tude of rock fractures have received some statistical treatment by Reiche [30]; Chayes [5]; and Pincus [28].

Inasmuch as textural and compositional features lend themselves well to laboratory study, a vast amount of statistical data has accumulated on these properties. Table 1 defines some textural and compositional properties of sediments, indicates those which have been quantified, and suggests the nature of the distributions obtained in each. Most work has been done on particle size distribution and mineral com-

TABLE 1

STATISTICAL ASPECTS OF SEDIMENTARY TEXTURES, COMPOSITION, AND MASS PROPERTIES

Property	Definition	Frequency Distributions*	
		Within single samples	Among closely spaced samples
Particle Size	Expressed as sieve mesh, intercepts, or in terms of settling velocity.	Log normal	Log mean is normally distributed.
Particle Sphericity	Cube root of ratio between particle volume and volume of circumscribing sphere.	Normal	Mean sphericity is normally distributed.
Particle Roundness	Ratio of average radii of edges to radius of circle inscribed in maximum projection plane.	Normal	Mean roundness is normally distributed.
Particle Surface Texture	Minute surface irregularities on particles. Definitions not quantified.		Percentage of frosted grains is normally distributed.
Particle Orientation	Orientation of particle axes or planes in space.	Normal or circular normal	Mean orientation is normally distributed.
Mineral Composition	Percentage composition of minerals present.	Discrete distributions. Binomial and Poisson(?) distributions.	Percentages of some minerals are normally distributed.
Porosity	Percentage of pore space in aggregate.	Normal	Normally distributed.
Permeability	Measure of ease of fluid flow through aggregates.	Log normal(?)	Log normal(?) distribution.
Natural Moisture Content	Percentage of moisture in freshly collected samples.	Normal	Normally distributed.

* Exceptions to the generalizations in these columns occur, but available data suggest that most distributions approach normalcy or log normalcy in their behavior.

position. Particle shape (sphericity and roundness) has been fairly extensively studied, whereas surface textures (such as frosted grains, striated grains, etc.) have hardly been approached statistically because the definitions cannot at present be operationally converted to numbers. The interested reader will find a discussion of methods and geological evaluation of the results in Krumbein and Pettijohn [18] and Pettijohn [27].

Aggregate properties (mass properties) of sediments depend on the associations of particles present in the deposit. Only three of a large number of aggregate properties are listed in the table. Some statistical work has been done on most mass properties, but in many instances it was confined to determination of mean values and degrees of spread.

Table 1 distinguishes between distributions of the variates within single samples, as against distributions of mean values from closely-spaced samples. Mass properties usually yield only a single value per sediment sample, although subsamples from a larger sample tend to distribute themselves as shown.

Information available for the last column of Table 1 is somewhat meager, although the kinds of distributions listed appear to apply. In many instances numerical characteristics of sedimentary phenomena show exponential rates of change when studied over long distances or in large areas. Frequency distributions of sample means taken over such larger areas sometimes are skewed, and may in some instances approach log normalcy. Although some percentage data are normally distributed as indicated, exceptions occur among rarer constituents.

Because size distributions and mineral data were among the earliest investigated, they are used here for illustration to indicate the growth of statistical methodology in sedimentation.

PARTICLE SIZE DISTRIBUTION OF SEDIMENTS

Sedimentary particles range in size from the order of 10^{-4} to 10^{4} mm. in diameter. Some sediments such as glacial till include this entire range; others have very restricted ranges of size, such as dune sand, which extends from about 0.1 to 1.0 mm. All workers with soils and sediments (geologists, soil scientists, engineers) realized early that some sort of geometric size scale is necessary to facilitate analysis and permit comparison of data. In geology the most widely used grade scale is based on the ratio 2. The reference value is 1.0 mm. and the scale extends in both directions, as 2, 4, 8 $\cdots$ mm., and $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ $\cdots$ mm.

Aside from technical problems of size analysis, not considered here,

early workers had the problem of statistical analysis of data arranged in unequal classes. It was felt that each size grade should have equal geometric significance (i.e., a change from 1 to 2 microns may be as important as a change from 1 to 2 mm.), and the uniformity or non-uniformity of the distribution should be expressible in such manner that fine or coarse sediments can be described in similar terms. These conditions were satisfied at an early date by the simple expedient of drawing histograms with equal width blocks for each geometric grade size, regardless of the absolute value of the class limits. The earliest such histograms known to the writer were used by Udden in 1898 [35].

There is no clear evidence that an implied log transformation was recognized at the time, although shortly after arithmetic cumulative curves were introduced in 1920, they were converted to their log equivalents by plotting them on semilog paper. These practices are still followed, inasmuch as nearly all histograms are shown with equal width blocks, and most cumulative curves are drawn on semilog paper for direct reading of median and quartiles. Log probability paper was used in the middle 1930's and in 1936 the writer [17] introduced a log transformation to facilitate conventional statistical analysis. The transformation is given by the relation $\phi = -\log_2 d$, where d is the diameter in mm. The minus sign was used for adjustment of graphical methods commonly used by geologists. The phi notation permitted direct application of moment analysis to size data and permitted definition of a normal phi curve described by the phi mean and phi standard deviation. This concept was extended to a phi Gram-Charlier series, and became the basis for graphic methods introduced by Otto [25] and extended by Inman [15]. Use of the phi mean and phi standard deviation instead of median and quartiles also permits convenient extensions of analysis to curve fitting, use of Chi square, applications of analysis of variance, etc.

Although many sediments approach log normalcy, the finer-grained sediments tend to be only partly symmetrized by the phi transformation. The writer experimented with other transformations to normalize these distributions, but on the whole it seems that most sediments can be described by the first four phi moments.

A problem of some interest in size analysis is the use of weight percentage frequency instead of number frequency. This question has been examined by Krumbein and Pettijohn [18], on the basis of earlier work by Hatch [14], who showed that if the number frequency distribution is log normal, the corresponding weight-frequency distribution is also

log normal, with the same log standard deviation, but with a systematic change in the log mean. In part, the problem appears to be one of convenience; for most analyses the number of grains per sample is very large and weighing is a more convenient means of expression. In coarse sediments where individual pebbles can be handled, and in microscopic examination of loose grains, number frequency is used.

TABLE 2

SEDIMENTARY APPLICATIONS OF STATISTICAL DATA

Application	Examples
1. Graphic presentation	Histograms, cumulative curves, frequency curves, scatter diagrams, etc.
2. Summarized description and comparison	Summaries of means, standard deviations, and other parameters; tests of relations among sediments.
3. Classification of sediments	Statistical data as a basis for textural and other groupings of sediments.
4. Study of sedimentary characteristics as functions of time and space	Graphs and maps showing systematic changes in sediments along streams, beaches, or over areas. Experimental design as a basis for studying population gradients and changes.
5. Use of statistical data in development and testing of dynamic theories of sediment behavior and variation	Comparison of field and laboratory data on sediment transportation and deposition with theories of sediment behavior derived in part from stage (4).

Frequency is always the dependent variable in size analysis and although the physical interpretation of the data may vary with the manner of expressing frequency, the geometrical significance of the statistical measures is the same regardless of the particular choice of frequency expression. The log transformation does not affect the relative area under any grade size block and hence is independent of the kind of frequency used.

The statistical data of particle size analysis have been used in various ways. Table 2 summarizes these uses, which in a broad way are the same as for other sciences. Most effort in the study of sediments has

been directed toward the first three categories. These descriptive uses have been necessary in the sense that the characteristics of sediments had first to be determined, and studies were needed of the areal variations in these characteristics over any given sedimentary environment. Once the range of values observed in nature was determined and some insight gained into the patterns of areal variation, it became possible to relate these observations to theory. In some instances theory precedes practice, and the observational data are used to test the theoretical structure.

MINERAL ANALYSIS OF SAND

The composition of sediments is expressed either in terms of chemical or mineralogical composition. Among sediments most extensively studied for mineral composition are sands, and in this section these sediments will be used as an example.

Most sand deposits are composed predominantly of quartz, but nearly all sands have small amounts of dark minerals which are important in sediment interpretation. These minerals have a greater specific gravity than quartz and are separated from it with heavy liquids. The heavy minerals, which are thus separated out, may comprise from 0.1 to 5 or 10 per cent of the sand.

In analyzing the heavy minerals, it is found that they may consist of few to a dozen or more species. A sample of several hundred to a thousand grains is studied under the microscope and the frequency of the species present is indicated as a number percentage. Inasmuch as there is no gradation among the mineral species, the distributions are multivariate with discretely measured variables.

The heavy mineral data are used partly to determine the kinds of source rocks which supplied the sediments, and partly to help decide whether two layers of sediment may be stratigraphically equivalent. Various statistical devices have been developed for these purposes.

In many early studies of heavy minerals, the relative abundance of species was indicated by such terms as "common," "rare," etc., although the use of number or percentage frequencies became common in the early 1920's. Problems quickly arose regarding the number of grains to be counted in order to avoid undue errors in estimating the rarer grain frequencies. Dryden [8] applied probable error theory to the problem in 1931 and concluded that about 300 grains should be counted. A second problem, investigated in 1935 by Dryden [9], concerned the comparison of heavy mineral suites among samples from

different strata to test the geological equivalency of the beds. His approach led to some discussion with Eisenhart [10] regarding the relative advantages of the correlation coefficient and the Chi-square test. Eisenhart's discussion clarified an important question in treating such data, and furnishes an excellent example of the contributions which statisticians can make to subject matter fields.

In 1944 Allen [1] reviewed the problem of applying statistical methods to mineralogical data. In 1949 he [2] applied the methods to a study of mineral variations in some Cretaceous deposits in England. Maps were presented showing areal variation in heavy mineral percentages, and scatter diagrams with regression lines showed relations among mineral composition, particle size, etc. Allen demonstrated that certain "patchy" occurrences of minerals were due to small scale nearly random processes, which however did not seriously affect the regional picture of mineral variation.

Allen's work, in common with other heavy mineral studies, was directed toward showing the *provenance* (place of origin and kind of parent rock) of the sediments and the main directions of material transport. Other studies are directed toward determination of the "stability" of minerals, i.e., their tendency to persist for long distances or times of transport. Various authors have explored this problem statistically; Pettijohn [26] showed that an order of mineral stability could be erected which indicates the relative persistence of heavy minerals due to resistance to abrasion, solution, or decomposition.

The statistical nature of heavy mineral distributions has not been given intensive treatment in the literature. Most sands consist of a large preponderance of quartz (plus detrital chert, feldspar, and some other "light minerals") and small amounts of heavy minerals. All the minerals have discrete number frequency distributions. The more abundant ones follow the binomial law, and the rarer ones appear to be Poisson distributions.[1] Many minerals show a normal distribution of mean percentages in closely spaced samples, as suggested in Table 1.

Observed and expected distributions of pebble frequencies in Lake Michigan beach gravel are shown in Table 3, based on studies by the writer. One hundred samples of 10 pebbles each were drawn at random from a small beach area, and the occurrences per sample of each rock type were noted. The gravel consists of about 50 per cent limestone, 35 per cent chert, 10 per cent basalt, and 5 per cent granite. The limestone

[1] I. W. Burr, in his oral discussion of this paper, suggested that the data of rarer minerals may be binomial distributions of low probability rather than Poisson. This point is considered below.

and chert values agree with their expected binomial distributions with P of the order of 0.50 by a Chi-square test. The limestone data are shown in the left of Table 3. The granite and basalt data agree with expected Poisson distributions, again with P exceeding 0.50. Following Burr's suggestion, however, the granite data were also compared with a binomal of low probability, and the Chi-square test yielded a P of about the same value. The right hand part of Table 3 shows the granite

TABLE 3

LITHOLOGIC COMPOSITION OF LAKE MICHIGAN BEACH PEBBLES

Number of Occurrences per Sub-sample	Limestone		Granite		
	Observed	Expected	Observed	Poisson Expected	Binomial Expected
0	0	0.1	58	58.8	59.9
1	1	1.0	33	31.2	31.5
2	6	4.4	7	8.3	7.5
3	7	11.7	2	1.5	1.0
4	23	20.5	0	0.2	0.1
5	26	24.6			
6	21	20.5			
7	12	11.7			
8	3	4.4			
9	1	1.0			
10	0	0.1			
Total Number of Subsamples	100	100.0	100	100.0	100.0

data and expected values for both Poisson and low-probability binomial distributions.

The writer has not explored these implications fully, but Burr's suggestion opens a fresh viewpoint in the study of sediment composition. It furnishes an additional example of the contributions which statisticians can make to subject matter investigations. Geologically the question is important because it affects interpretation of mineral data from samples collected along lines of natural transport, as in streams. Abundant minerals of low stability, which display binomial distributions near their source, become depleted as solution and abrasion act on them during transport. Do the binomial distributions merely change

by decrease of p, or is there some point at which the binomial law gives way to a Poisson law?

It is evident from the preceding discussion that much remains to be learned about the multivariate distributions of minerals in sedimentary deposits. Experimental design has entered the field only slightly, and there is ample opportunity for fundamental research in this aspect of sedimentary petrology.

RELATIONS AMONG SEDIMENTARY PROPERTIES

A wide variety of sedimentary techniques has been used in studying relations among properties of sedimentary rocks. Scatter diagrams and correlation coefficients have been widely used in testing relations among particle size, shape, composition, and the like [18, Chapter 9]. Many sedimentary characteristics vary exponentially with distance from source, and relations between the properties themselves independent of distance are commonly power functions as may be expected.

In the study of ancient sediments, the conditions of origin must be wholly inferred from relations among sediment properties and enclosed fossil organisms. In part, present day sediments are investigated to provide some basis for such interpretation. Comparisons of sediments formed under known conditions also yield data on the extent to which similar sediments may be produced by different environments. Beach sand and dune sand provide an excellent example, inasmuch as the responsible agents are waves and currents on one hand and wind on the other. In many instances the dune sand is derived from beach sand by selective wind transport, so that a continuous gradation is commonly discernible between the two. The question whether a single unknown sample is either beach or dune sand usually cannot be answered, although some separation of the populations can be effected with a group of samples.

Tests of relations among sedimentary properties commonly do not include evaluation of experimental and other errors. Chief reliance has been placed on the apparent spread or concentration of data in scatter diagrams. Where high correlation exists, and the data are not confused by use of inappropriate ratios, these analyses are probably sound. The problem of ratios and rates in studying geological data is one that requires further study. Some variables (such as grain sphericity and roundness) are defined in terms of ratios and yet show essentially normal distributions. In some instances the use of ratios may disguise relations among raw data as Chayes [4] pointed out in 1949.

Many geological studies suffer somewhat from failure to relate sample statistics to the corresponding population parameters. In many instances the sample statistics are used directly without determining standard errors or without applying tests for normalcy. Fortunately, most sediment properties have sufficient variation areally so that the experimental errors do not unduly cloud the relations. Many sedimentary samples display slightly skewed distributions, and some are highly skewed. A large number of the former show a reasonable probability of coming from normal populations (*P* commonly is greater than 0.10 in Chi-square tests). In part the generalization of Table 1 is based on such tests. Many samples are sufficiently skewed to reduce the likelihood that they came from normal populations, but relatively little has been done to investigate the geological conditions which may produce skewness. Similarly, the peakedness of some sedimentary distributions is greater than normal populations show. As with skewness, little has been done on this aspect, although there are suggestions that long-continued movement and agitation of sediments by geological agents may produce highly peaked symmetrical size frequency curves.

A large number of measured characteristics of sediments show bimodal or polymodal tendencies, which seem mainly to be the result of mixing effects under rapidly-shifting environmental conditions or of composite sampling which includes more than a single population. Considering that some sedimentary laminae may be a millimeter or less thick, the mechanical process of obtaining an unmixed sample may present a serious problem. Otto [24] introduced the concept of a "sedimentation unit" in 1938 as a basis for critical sampling of thin units.

SAMPLING PROBLEMS IN SEDIMENTATION

The foregoing brief discussions of size and mineral data indicate some of the main lines of statistical development in sedimentation. Most other sedimentary attributes were quantified after size analysis had become established, so that they were able to profit by the statistical experience of earlier work.

Several problems only partly solved are shared in common by all sedimentary fields. A principal one is that of sampling. It has been known for some time that the means of samples collected in a small area tend to be normally distributed for many sedimentary properties. If the samples from any one deposit are spread over a larger area, the distribution of sample means tends to develop a larger variance or to become skewed. In part, these changes are related to areal variations in

the population brought about by changes in the physical and chemical conditions of sedimentation over the large area. Maps which bring out these systematic changes can be made of the mean values or of the variances.

Two kinds of sampling problems commonly arise in sedimentary studies. One is concerned with the small-scale local variations of the sediment, and the other involves the large-scale or regional variations. These problems are met in oil exploration, for example. Oil occurs in sedimentary rocks and shows some relation to optimum sedimentary conditions. In studying potential oil-bearing areas, how far apart should samples be spaced to bring out the regional sedimentary trends? How closely should they be spaced to bring out the local departures from the regional trends? In part, oil occurs in areas which show anomalous variations from the regional picture. As some guide to the scale of thinking, regional maps may cover areas as great as 100,000 square miles. Regional samples from boreholes may be spaced about one per 500 square miles, and maximum close spacing is about 60 wells per square mile in thoroughly explored areas.

The study of recent sediments may be illustrated by the problem of sampling a beach 200 feet wide and several miles long. The deposits vary across the beach, along the beach, and with depth below the surface. The total volume of sediment may be of the order of 10^8 cubic feet, and the number of grains in the population may be of the order of 10^{16}. Along with the sand samples, data are to be collected on beach slopes, wave energy, strength of currents, etc., so that relations between sand characteristics and geological processes can be studied. How many samples should be taken, how should they be distributed over the beach, and how deeply should each sample penetrate the sand layers?

Solutions of such sampling problems have been largely empirical. If the upper layers are to be emphasized as being most nearly related to contemporary processes, closely spaced shallow samples commonly are collected along beach profiles spaced about ½ mile apart. An average of the samples along the profiles is used to characterize each ½-mile point along the beach. Presumably, the closely spaced samples compensate for local variations and the widely spaced averages bring out the sedimentary trends.

Designed experiments, which include the sample layout and specify the number and kinds of samples necessary for any given study, seem to offer one of the best approaches toward the sampling problem. As analysis of variance methods become more familiar to geologists, it is likely that planned experiments will dominate over the somewhat less

organized field studies which have been the rule in the past. Cochran's recent book, [7], published since this manuscript was written, carries many suggestions for stratified or systematic sampling plans which can be designed for particle populations.

Some studies have been directed toward evaluation of sampling and laboratory errors in size and mineral analysis. The writer [16] studied the probable error of sampling beach sands in 1934; more recently Griffiths [13] applied analysis of variance to the same data to sharpen the concepts. In 1937 Otto [23] applied Shewart's theory of control to the improvement of a splitter for obtaining subsamples from larger field samples.

Most error studies include evaluation of laboratory errors (sample splitting, sieving, weighing, etc.), as well as field variation from sample to sample. As an indication of magnitudes, it was found that the laboratory error in particle size analysis of beach sand was only about $\frac{1}{8}$ as large as the field sampling error (0.54 against 4.5 per cent) for the average diameter. In heavy mineral studies of the same samples the average laboratory error, assigned mainly to counting errors, was about equal to the field sampling error. Each was of the order of 10 per cent.

Griffiths and co-workers [11, 12, 31] applied two- and three-factor analysis of variance models to evaluation of sample, operator, and technique effects in grain orientation and porosity studies. The designs included evaluation of interactions. Perhaps the earliest analysis of variance study in the geological literature is that of Swineford and Swineford [33] published in 1946, in which a three-factor model was used to test the relative efficiency of sieve shaking equipment.

It is in the field of experimental design that some of the more important advances in statistical analysis of geological data will be made. Unlike many physical sciences, laboratory experimentation is not the main source of observational data in geology. Rather, the geologist must rely on field observations of natural processes and deposits for his data. Experimentation plays its part in studies of sand movement in water flumes and wind tunnels.

CONCLUDING REMARKS

The original notes for this paper were prepared during the autumn of 1952. It was apparent at the time that rapid developments could be expected in application of newer methods of statistical analysis to geological data. Important papers on analysis of variance had appeared in several fields [3, 6, 22], problems of sampling were being examined more closely, and workers were beginning to relate sample statistics to

population parameters [34]. The Earth Science Panel of the Committee on Statistics in the Physical Sciences of the A.S.A. had been organized and was bringing interested workers together. Early in 1953 a symposium of papers on statistics in geology was organized under R. L. Miller's supervision by the *Journal of Geology*. The November, 1953, and January, 1954 issues are devoted to the subject, and include several papers reviewing or extending applications of statistical analysis to sediments.

In some respects the last several years mark a resurgence of statistical interest in geology, directed toward more critical analysis of sedimentary data, and especially toward applications of experimental design. This is in contrast to earlier principal interest in developing and adapting techniques for description and classification of sediments, which reached its climax in late prewar years. A glance at the 1953 developments in the expanded bibliography [12, 13, 19, 20, 21, 29, 31] will indicate the directions of some of these later trends in sedimentary statistical analysis.

In the newer developments of statistical application geologists are becoming increasingly aware that progress in advanced statistical anlysis requires co-operation from statisticians. Opportunities for fuller collaboration between earth scientists and statisticians, such as are provided by the Committee on Statistics in the Physical Sciences, will do much to bring subject-matter and methodology groups together.

REFERENCES

[1] Allen, Percival, "Statistics in sedimentary petrology," *Nature*, 153 (1944), 71–77.

[2] Allen, Percival, "Wealdon petrology," *Quarterly Journal of the Geological Society*, 104 (1949), 257–321.

[3] Burma, B. H., "Studies in quantitative paleontology," *Journal of Paleontology*, 22 (1948), 725–61; 23 (1949), 95–103.

[4] Chayes, F., "On ratio correlation in petrography," *Journal of Geology*, 57 (1949), 239–54.

[5] Chayes, F., "Statistical analysis of two-dimensional fabric diagrams," in H. W. Fairbairn, *Structural Petrology of Deformed Rocks*, Cambridge, Massachusetts: Addison-Wesley Press, 1949, 297–307.

[6] Chayes, F., "The finer-grained calkalkaline granites of New England," *Journal of Geology*, 60 (1952), 207–54.

[7] Cochran, William G., *Sampling Techniques*, New York: John Wiley and Sons, 1953.

[8] Dryden, Lincoln, "Accuracy in percentage representation of heavy mineral frequencies," *Proceedings of the National Academy of Sciences*, 17 (1931), 233–38.

[9] Dryden, Lincoln, "A statistical method for the comparison of heavy mineral suites," *American Journal of Science*, 229 (1935), 393–408.

[10] Eisenhart, Churchill, "A test for the significance of lithological variations," *Journal of Sedimentary Petrology*, 5 (1935), 137–45.

[11] Griffiths, J. C., and Rosenfeld, M. A. "Progress in measurement of grain orientation in Bradford sand," *Producers Monthly* 15 (1951), 24–36.

[12] Griffiths, J. C., and Rosenfeld, M. A., "A further test of dimensional orientation of quartz grains in Bradford sand," *American Journal of Science*, 251 (1953), 192–214.

[13] Griffiths, J. C., "Estimation of error in grain size analysis," *Journal of Sedimentary Petrology*, 23 (1953), 75–84.

[14] Hatch, T., "Determination of average particle size from the screen analysis of non-uniform particulate substances," *Journal of the Franklin Institute*, 215 (1933), 27–37.

[15] Inman, Douglas, "Measures for describing the size distributions of sediments," *Journal of Sedimentary Petrology*, 22 (1952), 125–45.

[16] Krumbein, W. C., "The probable error of sampling sediments for mechanical analysis," *American Journal of Science*, 227 (1934), 204–14.

[17] Krumbein, W. C., "Applications of logarithmic moments to size frequency distributions of sediments," *Journal of Sedimentary Petrology*, 6 (1936), 36–47.

[18] Krumbein, W. C., and Pettijohn, F. J., *Manual of Sedimentary Petrography*. New York: Appleton-Century, 1938.

[19] Krumbein, W. C., and Miller, R. L., "Design of experiments for statistical analysis of geological data," *Journal of Geology*, 61 (1953), 510–32.

[20] Krumbein, W. C., "Statistical designs for sampling beach sand," *Transactions of the American Geophysical Union*, 34 (1953), 857–68.

[21] Manning, John C., "Application of statistical estimation and hypothesis testing to geological data," *Journal of Geology*, 61 (1953), 544–56.

[22] Miller, R. L., "An application of the analysis of variance to paleontology," *Journal of Paleontology*, 23 (1949), 635–40.

[23] Otto, George H., "The use of statistical methods in effecting improvements on a Jones sample splitter," *Journal of Sedimentary Petrology*, 7 (1937), 110–32.

[24] Otto, George H., "The sedimentation unit and its use in field sampling," *Journal of Geology*, 46 (1938), 569–82.

[25] Otto, George H., "A modified logarithmic probability graph for interpretation of mechanical analyses of sediments," *Journal of Sedimentary Petrology*, 9 (1939), 62–76.

[26] Pettijohn, F. J., "Persistence of heavy minerals and geologic age," *Journal of Geology*, 49 (1941), 610–25.

[27] Pettijohn, F. J., *Sedimentary Rocks*. New York: Harper and Brothers, 1949.

[28] Pincus, Howard J., "Statistical methods applied to the study of rock fractures," *Bulletin of the Geological Society of America*, 62 (1951), 81–130.

[29] Pincus, Howard J., "The analysis of aggregates of orientation data in earth sciences," *Journal of Geology*, 61 (1953), 482–509.

[30] Reiche, Parry, "An analysis of cross-lamination in the Coconino sandstone," *Journal of Geology*, 46 (1938), 905–32.

[31] Rosenfeld, M. A., and Griffiths, J. C., "An experimental test of visual comparison technique in estimating two dimensional sphericity and roundness of quartz grains," *American Journal of Science*, 251 (1953), 553–85.

[32] Simpson, G. G., and Roe, A., *Quantitative Zoology*. New York: McGraw-Hill Book Company, 1939.

[33] Swineford, Ada, and Swineford, Frances, "A comparison of three sieve shakers," *Journal of Sedimentary Petrology*, 16 (1946), 3–13.

[34] Strahler, Arthur N., "Equilibrium theory of erosional slopes approached by frequency distribution analysis," *American Journal of Science*, 248 (1950), 673–96; 800–14.

[35] Udden, Johan August, "The mechanical composition of wind deposits," *Augustana Library Publications No. 1*, 1898.

RELATIONSHIP BETWEEN AN INDEX OF HOUSE PRICES AND BUILDING COSTS*

DAVID M. BLANK
Columbia University

DEFLATION of residential wealth and residential construction expenditure estimates to constant dollar levels in principle requires the use of a price index of residential construction. However, no national market price index covering a reasonably long period of time exists, although house price indexes have been constructed for several cities, usually covering a relatively few years.[1] Consequently, a construction cost index is typically used as a substitute, on the view that the movement of such an index is a resonable reflection of changes in new house prices. This article attempts to assess the validity of this assumption and to judge the margins of error involved in employing a construction cost index as a deflator.

POSSIBLE DIVERGENCE BETWEEN COST AND PRICE INDEXES

It could be reasoned that significant short and long-term divergences might arise between a valid index of the market price of homes and indexes of construction cost. These divergences can be assigned to two causes: technical problems in defining and measuring construction costs and real deviations between new and old house prices.

Construction cost indexes usually exclude builders' profits and, often, overhead charges, or add a constant percentage to direct cost to cover these items. The apparently wide short-term variability in builders' profits thus permits significant differences to arise between the movement of prices of new homes and cost indexes. To the extent that there has been a secular movement of builders' profits and overhead costs which is not taken account of in construction cost indexes, even secular divergences may arise.

More importantly, technical problems inherent in devising construction cost indexes involve at least the possibility of deviations between

* Some of the material contained in this article is derived from a study of *Capital Formation in Residential Real Estate* by Leo Grebler, David M. Blank, and Louis Winnick, which is to be published by the National Bureau of Economic Research.

[1] For example, Toledo—William Hoad, *Real Estate Prices* (unpublished doctoral dissertation, University of Michigan, 1942); Washington—unpublished data from the Housing and Home Finance Agency, quoted in Ernest M. Fisher, *Urban Real Estate Markets: Characteristics and Financing* (New York: National Bureau of Economic Research, 1951), p. 54; Ann Arbor—Herman Wyngarden, "An Index of Local Real Estate Prices," *Michigan Business Studies* (Ann. Arbor: University of Michigan, January 1927).

such indexes and a true price index of new homes. Most residential construction cost indexes apparently are derived as some form of weighted average of materials prices and wage rates. They differ in the number of materials and labor skills covered and in the degree to which the weights are based on specific, rather than generalized, types of construction. The weights are usually unchanged, or changed little, over the entire period covered. Such indexes suffer from several defects. One such defect results from the fact that indexes of this kind cannot take fully into account the changing importance or the relative price movements of new materials and equipment which have been added to the house over time. In addition, for early years there is a serious question as to whether actual prices and wage rates, rather than nominal prices and rates, have entered into such indexes. Finally, such indexes are unable to take into account changes in site productivity.

If these technical problems were solved, cost indexes would properly measure the changes in prices of *new* homes. However, discrepancies between such cost and price indexes and an index of *old* home prices could still arise. Because of the interconnection between the markets for new and old homes their price movements should be in close conformity at most times. Nevertheless, divergences could appear at the trough of the building cycle when the prices of existing homes may sink below the price at which new houses would be offered on the market if there were any building activity. Indeed, it is for this reason that construction volume sometimes declines to more or less negligible levels. Discrepancies could also appear in the upswing of the cycle during short periods when new construction lags behind the increase in demand for dwelling units; existing houses may command premiums at such times because of their immediate availability. At either cycle stage, the divergences may last as long as several years.

A NEW PRICE INDEX, 1890–1934

To test the differences in movement between a representative construction cost index and the prices of homes, a house price index for 1890–1934 was developed and compared with the cost index. The data for the price index were derived from the *Financial Survey of Urban Housing*[2] which presented financial data and other information for a sample of residential structures in 61 cities in 1934. Detailed information in the *Survey* is available only for 22 cities. The 22 cities are widely scattered geographically, with at least two cities representing each of

[2] *Financial Survey of Urban Housing* (Washington, D. C.: U.S. Department of Commerce, 1937).

the nine census divisions except the East South Central division which had only one city in the 22-city sample. One set of questions asked of each owner of a residential structure related to: (a) value of the property in 1934, (b) year of acquisition by the then-present owner, and (c) original cost to owner at time of acquisition. This information was summarized for each city and a table presented for each of the 22 cities, listing the number of properties included in the 1934 sample which were acquired in each year from 1890 to 1933, the total acquisition cost of properties acquired in each such year and the value of each group of such properties in 1934. Separate data for all owner-occupied and all tenant-occupied structures and for all single-family owner-occupied houses and all single-family tenant-occupied houses were presented, rather than over-all figures for all residential properties.

The data selected for analysis were those relating to single-family owner-occupied houses, on the view that this relatively homogeneous group which comprises the most important portion of the nonfarm housing stock would show a more consistent pattern than the other categories. The all owner-occupied category might have been a reasonable alternative, but was rejected because it was less homogeneous than the single-family owner-occupied category. The two tenant-occupied segments were rejected because they included too small a number of properties and because the all tenant-occupied group was too heterogeneous. The tenant- and owner-occupied data could not be combined as they were based on two separate samples and the size of the two samples did not reflect the proportion of owner-occupied and tenant-occupied properties in the respective cities.

A relative for each year was calculated for each city, based on the ratio of the total acquisition cost of the single-family owner-occupied houses acquired in each given year in a given city to their value in 1934. The median relative for each year was then determined.[3] This series of median relatives, based on 1934 values equal to 100, was converted to a 1929 base; the converted series is presented in Table 1.

The assumptions underlying the price index warrant clarification before any comparisons are drawn. It is assumed, first, that the acquisition cost estimates are reasonably accurate. In all likelihood, the esti-

[3] To determine the effect of the specific averaging procedure on the final results, a test was performed on the data for a single year in each of the four full decades covered. The relatives for each such year were combined in the form of the median, positional mean, unweighted arithmetic mean, unweighted geometric mean, and weighted arithmetic mean (in which the weights were the number of households in each city at the nearest censal year). The range of results in each year was relatively small, so that the simplest measure, the median, was used in the computations for the final series. Individual city relatives based on less than four properties were disregarded in the computation of the median.

mates of acquisition cost for properties acquired in the early years of the period studies have some margin of error. It is also assumed that the year of acquisition has been accurately reported; here again, there undoubtedly are significant error margins for the early years, with a tendency for respondents to report acquisitions in years which are multiples of five. Finally, it is assumed that the movement between median relatives of two successive years approximates the movement in prices of a single sample between the two years; it will be remembered that each relative, before conversion to a 1929 base, actually represents the

TABLE 1

UNADJUSTED PRICE INDEX OF ONE-FAMILY OWNER-OCCUPIED HOUSES, 22 CITIES, 1890–1934

(1929 = 100)*

Year	Index	Year	Index
1890	61.3	1915	71.7
1891	55.3	1916	78.5
1892	56.3	1917	80.1
1893	58.7	1918	85.2
1894	68.4	1919	93.7
1895	62.5	1920	102.7
1896	53.8	1921	100.4
1897	55.5	1922	101.8
1898	59.1	1923	103.3
1899	56.5	1924	103.5
1900	64.6	1925	108.9
1901	54.2	1926	104.5
1902	63.9	1927	100.6
1903	64.9	1928	102.1
1904	67.9	1929	100.0
1905	59.5	1930	95.7
1906	70.6	1931	87.9
1907	77.9	1932	78.7
1908	70.3	1933	75.7
1909	68.5	1934	77.9
1910	74.2		
1911	72.5		
1912	75.3		
1913	75.3		
1914	78.1		

* Yearly median of 22 city relatives, excluding those relatives based on 3 or less properties.

movement in prices of a separate sample between the given year and 1934.

The validity of the 1934 value estimate probably does not seriously affect the movement of the price index, except for the 1934 value itself. It would only affect this movement if the degree of underestimate or overestimate of value in 1934 were correlated with length of holding.

It should be pointed out that the constructed price index applies to both new and old houses. The relative for a given year relates the acquisition cost of properties purchased in that year to their value in 1934, regardless of whether the acquisition was of a new or an old structure. A cursory examination of the data indicates that somewhat more than one-half of the properties in the 1934 sample which were acquired in the 1890–99 decade were new houses; somewhat more than one-third in the 1900–09 decade were new houses; and somewhat more than one-fifth in the remaining years were new houses. It was suggested earlier that there should be no reason for any difference in the price movement of new and old houses, other than in periods of depressed building activity or for short periods during the upswing of the cycle when consumer ignorance and relative availability may play a role. At other times, the movement of prices of houses of varying age and quality should be roughly similar. And the price variations of new housing, once it has entered the housing stock, should be the same as the original stock, subject to the same differential depreciation rates as apply to an existing housing stock composed of structures of different ages.

The index in its present form is subject to two major offsetting biases, viz., value losses due to depreciation and obsolescence and value increments in the form of structural additions and alterations. The price relative for 1904, for example, before conversion to a 1929 base, measures the change in price of a given set of properties between 1904 and 1934; this change is affected by the thirty years of depreciation operating on these properties and is somewhat smaller than the change in price which would be measured if this group of properties in 1934 had the same age structure as they did in 1904. Conversely, any structural additions or alterations to the properties between time of acquisition and 1934 would tend to make the price rise larger between these two periods than the theoretically correct price movement.

It is generally accepted that value losses due to depreciation and obsolescence typically outweigh value gains due to additions and alterations.[4] Therefore, the present index must be biased downward as the net result of these two kinds of value change.

[4] See Leo Grebler, David M. Blank, and Louis Winnick, *Capital Formation in Residential Real Estate—Trends and Prospects* (National Bureau of Economic Research, forthcoming), Appendix E.

Further corroboration for this view is found in a comparison of two sets of house price indexes for Cleveland and Seattle (Table 2). One set of indexes comprises the series of relatives for these two cities, which, together with the relatives for the remaining 20 cities, provided the basis for calculating the 22-city price index. These indexes are

TABLE 2

HOUSE PRICE INDEXES, CLEVELAND AND SEATTLE
(1929 = 100.0)

Year	Cleveland		Seattle	
	Garfield-Hoad Price Index (1)	Price Index Underlying 22-City Index (2)	Garfield-Hoad Price Index (3)	Price Index Underlying 22-City Index (4)
1907	35.4	64.7		
1908	36.6	60.8		
1909	40.2	66.5	56.9	76.4
1910	43.9	59.1	58.8	74.4
1911	45.1	57.7	56.9	82.9
1912	46.3	62.0	64.7	73.6
1913	47.6	63.8	62.7	78.0
1914	50.0	72.2	64.7	86.9
1915	51.2	70.0	66.7	86.9
1916	53.7	71.0	64.7	77.7
1917	58.5	77.2	62.7	76.3
1918	67.1	89.7	66.7	82.1
1919	76.8	89.6	78.4	92.6
1920	86.8	104.7	88.2	95.7
1921	87.8	102.9	86.3	93.5
1922	91.5	104.6	99.8	88.3
1923	96.3	101.1	100.0	94.2
1924	100.0	113.1	117.6	96.7
1925	102.4	112.9	109.8	102.9
1926	103.7	114.5	107.8	98.0
1927	102.4	106.1	99.9	98.2
1928	101.2	111.0	102.0	99.6
1929	100.0	100.0	100.0	100.0
1930	95.1	94.3	88.2	92.5

Sources:

Columns 1 and 3—Index derived from 3-year moving averages of prices paid for new six-room frame house and lot. Garfield and Hoad, *op. cit.*

Columns 2 and 4—Index for prices of 1-family owner-occupied homes. Derived from data in *Financial Survey.* Index one of 22 underlying 22-city price index.

subject to the same biases for depreciation and additions as the 22-city index itself.

The second set of indexes were derived in such a manner as to exclude any such bias. They are based on three-year moving averages of prices paid for new owner-occupied single-family homes in Cleveland and Seattle, derived by Frank R. Garfield and William M. Hoad from special tabulations of unpublished data from the *Financial Survey of Urban Housing*.[5] From these tabulations Garfield and Hoad were able to compute average prices paid for new homes (including the lots underlying the structures) of specified types in each city in each year covered. The authors focus attention primarily on the price movement of five- and six-room frame houses, on the assumption that changes in the transaction-mix would affect the averages but little, since an analysis of the distribution of prices paid for various types of homes purchased in Cleveland in 1924 had indicated that these were relatively homogeneous types of structures. The series for six-room frame houses in each city, converted to indexes with a 1929 base, are given in Table 2.

The properties underlying the Garfield-Hoad indexes may have been subject to changes in size and quality of structures and in land ratios which would result in divergences between these indexes and a valid house price index. But such changes were probably severely limited in extent due to the stated homogeneity over time of the houses with regard to size and type of structure and construction, i.e., 6-room single-family frame houses. And the restriction of the data to new houses specifically excludes any biases due to depreciation, obsolescence, or additions and alterations.

A comparison between the two sets of indexes shows a significantly greater rise in the Garfield-Hoad indexes between the pre-World War I period and the late twenties than in the price indexes for Cleveland and Seattle underlying the 22-city price index. This difference is fully consistent with the existence of a downward bias in the 22-city index due to the effects of depreciation gross of additions and alterations.

A detailed examination of empirical data, undertaken elsewhere, suggests that the decline in value of single-family houses over the first 52 years of life, resulting from the net effect of depreciation and obsolescence on the one hand and additions and alterations on the other, approximates that resulting from a 1.2 per cent linear rate of depreciation.[6] Since the 22-city index is based on movements in the

---

[5] Frank R. Garfield and William M. Hoad, "Construction Costs and Real Property Values," *Journal of the American Statistical Association*, December 1937, pp. 643–53.

[6] Grebler, Blank, and Winnick, *loc. cit.* The data were derived from a special study by the Federal

prices of structures plus land, the depreciation correction for this index also requires a rate based on structures plus land. The relevant linear rate, derived from the same data, is about 1.0 per cent.

All studies of the decline in market value of houses as they age clearly indicate that a curvilinear rate of depreciation is more appropriate for residential structures than a linear rate. The compound rate of depreciation which yields about the same remaining value after 52 years as a 1.0 per cent linear rate, but which approximates more closely the path of declining value of residential structures as they age, is about $1\frac{3}{8}$ per cent. Accordingly, the 22-city index was corrected for a $1\frac{3}{8}$ per cent compound rate of depreciation. The series so calculated, after adjustment so that 1929 again equals 100, is presented in Table 3.

Generally speaking, the corrected price index shows an upward secular drift from 1890 to about 1916, a more rapid rise to 1920, a smaller rise to 1925, and a decline thereafter to 1933. Between 1890 and about 1925, short cycles of about four years in duration are discernible in the data, with peaks appearing in 1894, 1900, 1904, 1907, 1910, 1914, 1920, and 1925.[7]

PRICE INDEX COMPARED WITH CONSTRUCTION COST INDEX

No residential construction cost index covers the entire period from 1890 to 1934 but the Boeckh residential construction cost index, based on 20 cities, starts in 1910 and can be extrapolated back to 1890 in a customary fashion by the use of building materials and building wage rate indexes. The Boeckh index is one of the few adequate construction cost indexes available and is the only one aimed specifically at measuring changes in cost of construction of residential structures.[8]

The combined index is presented in Table 4. The construction cost

---

Housing Administration of a sample of single-family homes appraised by FHA in 1939; from William M. Hoad, "Real Estate Prices, A Study of Residential Real Estate in Lucas County, Ohio," unpublished doctoral dissertation, University of Michigan, Ann Arbor, 1942; and from Raymond Goldsmith's analysis in "A Perpetual Inventory of National Wealth," *Studies in Income and Wealth*, Vol. XIV, (New York: National Bureau of Economic Research, 1951) of data gathered by the *Financial Survey of Urban Housing*. It should be pointed out that a 1.2 per cent linear rate of depreciation for houses is significantly below the rates presented in Bulletin F and used by the Department of Commerce and most other investigators in this field, even after all adjustments are made for comparability.

[7] The short cycle in house prices approximates closely in length the short cycle found in building activity by Long. Clarence D. Long, Jr., *Building Cycles and the Theory of Investment* (Princeton, 1940), p. 104.

[8] E. H. Boeckh and Associates actually construct ten indexes for different types of structures, both residential and nonresidential, for various cities. Two of these indexes, for frame and for brick one- to six-family residential structures, for 20 cities have been combined by the several successive federal housing agencies into a single residential cost index which is used by the Department of Commerce in deriving the residential construction expenditure component of the deflated Gross National Product series. It is this index which is referred to in the text.

TABLE 3

PRICE INDEX OF ONE-FAMILY OWNER-OCCUPIED HOUSES, 22 CITIES, CORRECTED FOR 1⅜ PER CENT COMPOUND ANNUAL DEPRECIATION 1890–1934
(1929=100.0)

Year	Index	Year	Index
1890	36.0	1910	57.3
1891	32.9	1911	56.7
1892	34.0	1912	59.7
1893	35.9	1913	60.5
1894	42.4	1914	63.7
1895	39.0	1915	59.2
1896	34.3	1916	65.8
1897	35.9	1917	68.0
1898	38.7	1918	73.3
1899	37.5	1919	81.7
1900	43.5	1920	90.8
1901	37.0	1921	90.0
1902	42.4	1922	92.5
1903	45.5	1923	95.2
1904	48.3	1924	96.7
1905	42.9	1925	103.1
1906	51.6	1926	100.4
1907	57.7	1927	97.9
1908	52.8	1928	100.7
1909	52.3	1929	100.0
		1930	97.1
		1931	90.4
		1932	82.0
		1933	80.0
		1934	78.3

Source: Index, Table 1, corrected for 1⅜ per cent compound annual depreciation.

index for 1890–1934 and the corrected house price index (Table 3) for the same period are compared in Chart I. A comparison of the two indexes suggests two important conclusions with regard to the relationship between construction costs and house prices.

Except for the period 1916–1922,[9] the price index shows more short-run variability than the cost index. The latter is quite stable over the

[9] The cost index rises to a much sharper peak in 1920 than does the price index. This sharp rise in 1920 is found in all construction cost indexes and probably reflects a real difference in construction costs and prices in that year. It seems to have been a result of a unique set of supply and transportation difficulties in the winter and spring of 1920.

TABLE 4

RESIDENTIAL CONSTRUCTION COST INDEX, 1890–1934

(1929 = 100.0)

Year	Index	Year	Index
1890	39.2	1915	53.5
1891	37.9	1916	57.0
1892	36.8	1917	66.6
1893	36.7	1918	79.2
1894	35.4	1919	92.1
1895	34.9	1920	118.7
1896	35.1	1921	95.4
1897	34.4	1922	87.7
1898	35.9	1923	98.3
1899	38.5	1924	96.9
1900	40.6	1925	96.2
1901	40.1	1926	96.9
1902	41.5	1927	95.6
1903	43.0	1928	95.9
1904	42.5	1929	100.0
1905	44.5	1930	97.5
1906	48.9	1931	89.9
1907	51.1	1932	76.1
1908	49.5	1933	76.2
1909	51.4	1934	82.9
1910	53.2		
1911	52.5		
1912	53.8		
1913	51.9		
1914	52.2		

Sources:

1890–1906: 1907 value extrapolated by weighted average of an index of average wages per hour in the building trades and an index of building materials prices. Wage index from Department of Commerce and Labor, *Bulletin of the Bureau of Labor*, No. 77, July 1908; see *Historical Statistics*, p. 66. Price index from *Handbook of Labor Statistics*, 1941 edition, Vol. 1; see *Historical Statistics*, pp. 233–34. Weights—wages, 1.0; materials, 1.5. Weights derived from NHA analysis of housing costs; see *Housing Statistics Handbook*, p. 32.

1907–1909: 1910 value extrapolated by weighted average of an index of wage rates in the building trades and an index of building materials prices. Wage rate index from Bureau of Labor Statistics annual reports, *Union Wages and Hours in the Building Trades*; see *Historical Statistics*, p. 69. Price index from same source as 1890–1906. Weights same as above.

1910–1914: 1915 value extrapolated by Boeckh index of residential construction cost, as given in *Historical Statistics*, p. 172.

1915–1934: Boeckh residential construction cost index, as given in *Construction and Building Materials, Statistical Supplement*, May 1951, Department of Commerce, p. 40, converted to 1929 base.

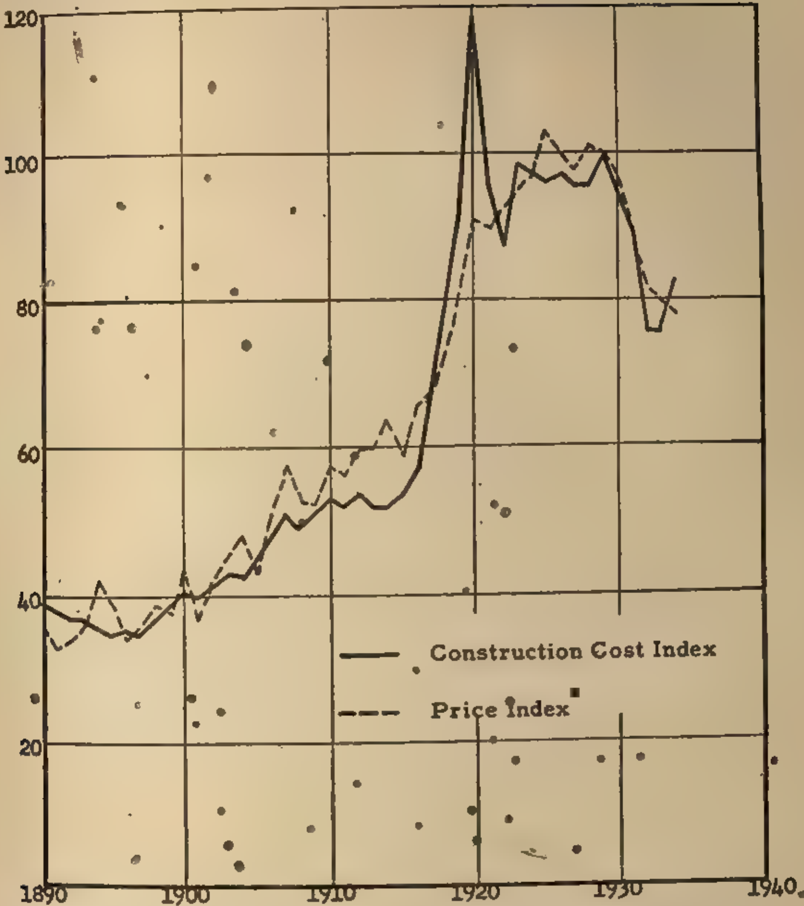


CHART I. Price index of single-family owner-occupied houses, corrected for depreciation, and residential construction cost index, 1890–1934. (1929 = 100.)

Source: Tables 3 and 4.

pre-1916 period, partly perhaps as a result of the way in which it is constructed, while the price index shows substantial fluctuations. Between 1905 and 1909, for example, the price index has a rise of more than 34 per cent and a fall of almost 10 per cent, as compared with the cost index which rises only 15 per cent between 1905 and 1907 and declines only 3 per cent between 1907 and 1908. The same relationship holds for the period after 1922; the price index falls 5 per cent between 1925 and 1927 while the cost index remains almost unchanged. In sum, it seems reasonable to conclude that in most periods the market

price of homes fluctuates more widely over the short run than do construction costs as measured by standard construction cost indexes. As a result, the annual movements of any construction series deflated by a construction cost index are subject to some margin of error.

But equally important is the fact that the long-run movement of the two indexes is remarkably similar. Thus, the construction cost index in 1921–1929 is about 245 per cent of its level in 1895–1905; the corrected price index in 1921–1929 is about 241 per cent of its level in 1895–1905. It must be remembered that the price data and depreciation data underlying the corrected price index are derived from independent sources and that both are completely independent of the cost data underlying the construction cost index. In view of this independence of derivation, the almost identical long-run movement of the two series over four and a half decades argues strongly that the construction cost index measures with quite reasonable accuracy the secular movement of house prices.[10]

CONCLUSIONS

The 22-city price index and the construction cost index show significant short-term divergences. These suggest that market prices of homes fluctuate more widely than construction costs, the difference in rise or fall perhaps amounting to as much as 10 per cent in a period of several years. For short-term analysis, then, some margins of error are involved in using the cost index as an approximation of a price index.

With regard to long-term movements, however, the construction cost index conforms very closely to the price index, corrected for depreciation. It would appear, therefore, that for long-term analysis the margin of error involved in using the cost index as an approximation of a price index cannot be very great.

[10] Only if there were major increases in site productivity not reflected in the construction cost index might this view be questioned. Although data on this question are extremely scanty, there is some evidence that the cost index is not subject to major error on this score, partly because the index does not fully reflect the historic increases in labor cost and partly because gains in site productivity of *residential* construction have been relatively limited, except perhaps in the very recent past. See Miles Colean and Robinson Newcomb, *Stabilizing Construction: The Record and Potential* (New York: McGraw-Hill, 1952), pp. 69–74, 247–248. Also Grebler, Blank, and Winnick, *op. cit.*, Appendix C.

CYCLES IN THE BALANCE OF PAYMENTS*

SOLOMON FABRICANT
New York University

LOOKING back at the vast literature on international trade theory, Jacob Viner noted in 1937 that the older discussions contained "only scattered and incidental references to the repercussions on the international mechanism of cyclical fluctuations in business activity" (*Studies in the Theory of International Trade*, p. 432). But he observed also that "within the last few years" the question was being "more seriously tackled;" and, as we now know, at the very time he wrote this spark of energy was already being blown into flame by the lively breath of Keynes's *General Theory*. In the past decade and a half, however, little of this energy has been devoted to the "inductive spadework on the international aspects of business fluctuations" that Viner rightly felt was one of the steps necessary in a fruitful attempt to "incorporate cycle theory into the theory of international trade . . . or to apply international trade theory to cycle theory."

The statistical analysis by Chang is therefore one of the exceptions. Indeed, it is the only sustained attempt to determine in some systematic fashion the characteristic cyclical behavior of the several items in the balance of payments of a variety of countries that has been published so far. Such enterprise merits applause—and in this case even wonder, for Mr. Chang, apparently working without statistical assistance, must have spent his days and his nights with the calculations.

Necessary perspective on Mr. Chang's work is provided if we start by asking ourselves what "inductive spadework on the international aspects of business fluctuations" should attempt to uncover. Suppose—let us dream—that we have unlimited time, money, and data. For any country, then, we want to know how each of the several items in its balance of payments fluctuates, and how each behaves during domestic and foreign business cycles. How well do changes in exports, for example, conform to changes in domestic business conditions? Do turns in exports usually lead, coincide with, or lag behind, turns in domestic business, or is their timing irregular; what is the average amplitude of fluctuation of exports, and the variation about this average, during domestic business cycles; do exports usually rise most vigorously during the early stages of domestic business expansion or during the later stages, or is there no systematic difference; does "quantum" of exports

* A review article of *Cyclical Movements in the Balance of Payments*, by Tse Chun Chang. Cambridge (England): Cambridge University Press. 1951. Pp. x, 224. $3.75.

fluctuate more or less than price of exports, and price of exports more or less than price of imports and other goods? Are there any systematic differences between the cyclical behavior of exports of raw materials and those of fabricated products, and if so, what are the relative weights of the several classes of goods, and how have long-run changes in these weights affected the cyclical behavior of the total? And we ask, also, if the behavior of exports seems about the same whether the domestic cycle is short or long, mild or severe; and if there are differences in cyclical behavior systematically related to the rate of growth of the economy or its several parts, or to other secular or structural changes, e.g., in tariff barriers and monetary standards. Then, looking over a country's borders at business conditions abroad, we consider how the latter are related to domestic conditions and to fluctuations in the items in its balance of payments. Having thus studied cycles in the balance of payments of each country—or an adequate sample of countries—we would go on to see whether the countries fall into homogeneous groups, each with its characteristic type of balance of payments fluctuation: whether, for example, "industrial countries" differ from "raw-material producers," or industrial countries heavily engaged in export production from those in which production for export markets is relatively small. And we would search also for similarities among countries with respect to secular change in behavior.

These questions would be put differently and in different order by a person with other tastes and theoretical predilections, but in one form or another they—and many other factual questions—would be in the list of every economist earnestly seeking light on the international aspects of business fluctuations.

To answer these questions—remember, we are dreaming—we would use many rather long statistical series, and most of these would be on a monthly or quarterly basis. We would use a statistical apparatus that enabled us to study the shape of individual cycles in each series, as well as averages. We would relate developments in each country to those in countries closely tied to it by trade and finance, as well as to those in the rest of the world as a whole.

It will be no surprise to the reader that Mr. Chang has answered few of our questions. Apart from the obvious reasons, others appear as we take stock of what he did.

For each of six countries (Britain, the United States, Sweden, Australia, Chile, and Canada) Chang took the annual figures for each major item in the balance of payments (except the net gold and capital flow) during the period 1924–1938, eliminated trends, and determined—one at a time—the linear logarithmic equation relating the fluctua-

tions in each series to those in the presumed causal factors. Reading these equations, he had the percentage by which each item in the balance of payments changes on the average with a one per cent change in each of the "independent" factors. These factors were then related by a similar process to real "world income" (more exactly, the real income of the rest of the world). Simple substitutions in the equations gave him the percentage change in each item which could be expected if world income rose or fell by one per cent. Applying these percentage changes to the figures in a base period, he calculated the corresponding absolute amount of change in each item. The net difference among them gave him the absolute change that could be expected in net capital exports, including gold and the residual error. The various equations, and a tabulation of the absolute changes to be expected when real world income rises or falls by one per cent, constitute his statistical summary of the cyclical fluctuations in the balance of payments of each country.

The results of this process of summarization may be illustrated by the American figures. The following table gives the regression coefficients (i.e., the elasticities):

Item	R[a]	Elasticity with respect to US Real Income	World Real Income	US Money Income	World Money Income	US Relative Import Price[b]	US Relative Export Price[c]
United States							
Quantity of imports	.91	1.27				−.97	
Quantity of exports	.97		2.92				−.43
Import Price	.91	1.38					
Export Price	.93	.94					
Interest receipts	.84				1.40		
Interest payments	.77			1.13			
Net tourist payments[d]	.90			1.22			
Net immigrants' remittances	.71			.41			
Net shipping payments	.69			1.14			
Real income	*		2.15				
Cost of living	*			.67			
Money income	*	1.36					
Competitors' price	*		1.08				
World							
Money income (in U.S. dollars)	*		1.75				

* Not given.

[a] Coefficient of multiple correlation.

[b] U.S. import price (with tariff) divided by U.S. cost of living.

[c] U.S. export price divided by U.S. competitors' export price.

[d] "1931 and 1932 are excluded because of the abnormal influence of world exchange depreciation." (p. 149)

Chang derives the "cyclical pattern of American income account," expressed in millions of dollars of change, by applying these elasticities to average values in the base period 1925–1926. For a one per cent rise in world real income (and a corresponding 2.15 per cent change in U. S. real income) he has:

Imports	−173.3
Exports	+195.0
Interest receipts	+ 18.9
Interest payments	− 7.9
Other current items	− 14.9
Net change in balance	+ 17.8

(All items rise; as usual, the minus signs represent debit items.)

What we get out of this for each of the six countries is the usual direction and average amplitude of fluctuation in each major item in its balance of payments relative to a given movement in the "world cycle" (or, if one wishes, in the country's own real income), and some notion of the separate shares of price and quantity variation in import and export value change.

In order to be able to say something about "typical" differences between raw-material producers and industrialized countries, Chang supplements these calculations for the sample of six countries with a less detailed examination of the data for another 15 or 16 countries. For these countries he determines the first two of the equations given in the above list for the United States, and thus obtains the elasticities of quantity of imports with respect to real domestic income and relative price (his Table 4) and of quantity of exports with respect to real world income and relative price (Table 6). All told, then, he has the factors determining (or associated with) the imports of 21 and the exports of 22 countries, with 19 common to both lists. Again, the period is 1924–1938, the data are annual, and the trends have been eliminated.

Chang finds, with respect to imports, that the price factor is of minor importance: all but one of the elasticities are below unity, and for 13 countries they are below .5, neglecting the minus sign. These findings tell us something about the extent to which price fluctuations are associated with quantity fluctuations (income constant). Chang's reading of them as demand elasticities in the neoclassical sense, however, raises questions of the sort that troubled economists in interpreting statistical demand curves during the 1920's. Indeed, Chang's interpretation of his results—reproduced from his earlier published articles—has already been criticized (see, for example, Guy Orcutt's article in the May 1950 *Review of Economics and Statistics*).

Chang concludes that the important demand factor is income: all but one of the income elasticities are above unity, with those of "industrialized" countries lower than those of agricultural and mining countries. That is, "given a uniform economic expansion or contraction all over the world, the quantity of imports of the industrial countries tends to fluctuate less than that of the world average; and, that of the agricultural [and mining] countries tends to fluctuate more" (p. 43).

With respect to exports, the price elasticity (again ignoring the minus sign) is above unity for 4 countries, between .5 and 1.0 for 9, and below .5 for the remaining 9. Income elasticities are above unity for virtually all the industrial and mineral producing countries (Canada is classed here as mining; in Table 4 it was classed as agricultural); they are below unity for all the agricultural countries. "Looking at the results broadly, we find that Table 6 depicts a situation reverse to that of Table 4. The countries whose import income elasticity is less than that of the world as a whole are those whose export income elasticity is greater than that of the world as a whole; and conversely. Or, speaking more generally, for the former cases, the import income elasticity tends to be smaller than the export income elasticity; whereas, for the latter cases, the export income elasticity tends to be smaller than the import income elasticity" (p. 51).

Chang is rather careless here, for he is comparing the import elasticity of a country with respect to its own income, and its export elasticity with respect to world income. It is something of a jump to imply here, and assert explicitly later (p. 170), that "the difference in the magnitude of import and export income elasticity of all the agricultural countries tends to result in a large and unfavorable change in relative quantities in prosperity." Except for the six countries studied in detail Chang presents no data showing how the income of individual countries fluctuates in relation to world income. However, the statement about change in relative quantites might be warranted. Agricultural output in the United States (and therefore presumably also purchases of such output by domestic "industry") fluctuates within a narrower range than does mining and manufacturing output (and therefore, presumably, purchases of such products by farmers). This is probably true of trade within other countries, and possibly of international trade as well.

Chang's elasticity calculations relate to quantities, not values. Except for the six countries he provides no evidence on the cyclical fluctuations in terms of trade. However, it is pretty clear, again from information outside of Chang's book, that the terms of trade change in

favor of agricultural communities when business improves. Further, Chang shows (for eleven agricultural countries combined, p. 169) that the balance of merchandise trade was negative during 1924–1930 and positive during 1931–1938. Chang is led to assert, therefore, that change in relative quantities tends to more than offset relative price change (p. 170).

The picture for industrial countries is the opposite, of course. In their case, the merchandise export balance rises as world business improves and falls as world business worsens.

In the case of mining countries, Chang believes the change in relative quantities to be small; but the change in relative prices to be great and in favor of these countries as world business expands. Therefore their merchandise export balance tends to behave like that of industrial countries.

If all this is true it means (ignoring minor items in the balance of payments) that net capital export by the industrial and mining countries tends to be greater (or net capital import smaller), during world prosperity than during world depression, with the reverse for agricultural countries. These generalizations about changes in capital flows and associated changes in trade quantities, prices, and values during fluctuations in world business constitute Chang's major findings.

It is clear that the scope of Chang's results is narrow and that we are far short of having all the answers to the questions listed earlier. But no investigator working alone could have gotten very far in filling that bill.

Another criticism, however, must be made of the adequacy of his results. There, too, the conclusion is clear. We are not sure of the answers that Chang has given us. And the grounds for our doubts are much the same as those that inevitably restrict the scope of his results.

One reason is the limited range of the data analyzed. Whatever apparatus is applied in the analysis, it cannot be expected to extract from short series of annual data reliable answers to our questions. It might perhaps be argued that the world economy of pre-1914 days was so different from that of the inter-war period, that we must make shift with the 1924–1938 data if our concern is with the workings of the economic world of that period; but this cannot be determined *a priori*, and I do not believe that it jibes with the known facts. Chang should not and need not have confined himself to the period 1924–1938. Nor should he have restricted himself to annual data, for annual data suppress too many significant features of cyclical change. Limits on his time and energy would have forced him to be less ambitious in other

directions, but I suspect that he chose the inferior alternative. Study of a longer period might also have forestalled some of the very serious questions that arise about his trend eliminations. Chang tells us practically nothing about them. Since we know that his period of fifteen years is short, and that it includes an exceptionally long cycle, we tend to imagine the worst about the adequacy of his trends.

Even the short series might have yielded more than Chang squeezed out of them had he used some other apparatus of analysis. In effect, he made up a scatter diagram on a double-log scale, plotting (say) exports against income, drew a regression line through the fifteen points in the scatter, read off the slope of the line, and discarded the basic data and the diagram. The regression coefficient (i.e., the elasticity) hardly tells all that might be learned from the basic data or even the chart. For example, is the slope greatly influenced by the extreme points; that is, is the slope a reflection mainly of a single large cycle—that of the 1930's—or does it fairly reflect all the cycles including the (two) smaller ones? Chang provides enough scattered information to raise some serious doubts about this, but no systematic examination is made nor is enough information given to enable the reader to undertake it for himself.

When three-variable scatters are used, another question arises. Multiple correlation coefficients are usually given by Chang. However, the standard errors of the regression coefficients are conspicuous by their absence, and nothing is said about intercorrelation between factors. This is a serious deficiency in a context in which the problem of multicollinearity is important.

There is a question also about the use of straight lines on double-log paper. Do they always tell the story? They could not, for example, express adequately the relation between an asymmetrical cycle in one series and a symmetrical cycle in another.

The reader will realize that I am suggesting the use of some such apparatus as Mitchell's in studying cycles; but was not this apparatus designed by an expert for the very purpose?

Chang assumes that a rather tightly-knit world economy existed in the inter-war period. This is a basic point, for his major finding is in terms of a world cycle. He notes at one place (p. 15) that the timing and intensity of cyclical fluctuation in the different countries are not the same, but on the next page dismisses this as unimportant. He seems to have been a victim of his choice of data and method of analysis, which led him to assume the world-wide diffusion of the depression of the 1930's to be characteristic of cycles generally. (A. F. Burns points

out that "after 1919 the business cycles of different countries tended to drift apart, though practically all shared in the catastrophic contraction of 1929–32;" see *Papers and Proceedings, American Economic Review*, May 1949, p. 82.) Careful examination even of that episode might have raised some doubts in his mind about the assumption of a closely integrated world economy. By failing to present data for more than the six countries on how the income of individual countries fluctuates in relation to world income, he fails to provide the basis for such an examination, and fails to present convincing evidence for his assertion that the "trade cycle is a world-wide phenomenon" (p. 220). Indeed, we know that the peak even before the contraction of the 1930's came at different times (even on an annual basis—see Thorp's business annals for 1926–1931, *News-Bulletin* of the National Bureau of Economic Research, Sept. 1932); and Chang's equations for the six countries suggest how greatly contractions have varied in severity (the "elasticity" of national real income with respect to world real income ranged from .58 for the U. K. to 2.15 for the U. S.).

As part and parcel of the above assumption Chang assumes that there is a single world market for all export countries and that world income is the dominant demand factor with respect to each. However, the "world market" is a group of markets, closely interrelated for some commodities, loosely for others: recall A. J. Brown's observations (in Chapter VI of his *Applied Economics*), and Chang himself notes that countries have "exclusive" markets and that "world market is an ambiguous notion" (pp. 53, 70). Since the incomes of the various importing countries do not in fact fluctuate identically, we cannot expect that all exporting countries will be confronted with the same demand conditions. The aggregate of world income is therefore hardly an appropriate measure of the strength of the demand confronting any individual country. But even if there were a world market, Chang's simple aggregate of national incomes could not be an adequate measure because it makes no allowance for international differences in import propensities.

Chang also ignores an important factor on the supply side. Agricultural output is influenced greatly by the weather. Because this varies over the surface of the earth, individual countries here and there will enjoy a bountiful harvest or suffer a drought, although world agricultural output will be fairly steady. The value of a country's agricultural exports might therefore be high when that of other countries is low, low when that of other countries is high. (And this in turn, by influencing domestic business conditions outside the agricultural sphere [see

Mitchell's *What Happens during Business Cycles*, p. 58, footnote, and his earlier 1913 report], will contribute to diversity of national income experience among the countries of the world.)

There are some questions, finally, about the accuracy of Chang's basic data, as well as the combinations he made of them. Students of income statistics will wonder about the adequacy of the real income estimates for the score of countries covered. Others will raise their eyebrows at some of the series on quantum and price of imports and exports. Chang mentions sources but does not go sufficiently into the details of the construction of the estimates or their adequacy for his purposes; nor does he indicate what change in his findings might result were use made of alternative estimates, where these are available.

We are left, at the end, with doubts not only about the applicability of Chang's findings for 1924–1938 to other periods, but also about their adequacy for the period he covered.

The chapter in Viner's book to which reference was made above opened with a statement unearthed by him from a work published in 1857: "Many writers have perplexed themselves and their readers by founding theories on exceptional circumstances. Others have been led astray by statistics—the characteristic form of modern research." Chang's book illustrates both dangers. Yet we should not forget that his study is the first attempt at a systematic survey of cyclical movements in the balance of payments of a wide variety of countries. His energy—even his boldness—in grappling with stubborn facts sets us an example. While we cannot follow in his footsteps all the way, we may discover the right path more easily because of his pioneering efforts.

DEMAND ANALYSIS*

H. S. Houthakker
University of Chicago

In a field where papers are plentiful but books are rare the appearance of a monograph by one of the leaders in its development raises high expectations. Professor Wold is perhaps best known as a mathematical statistician, but he has also made valuable contributions to economic theory [15] and his earlier empirical study on the demand for farm products [14] would no doubt be generally recognized as a classic if its language and time of publication had not curtailed its circulation. For the present work he associated himself with Mr. Jureén, a government statistician; he has also drawn upon the support of a number of other distinguished collaborators in Sweden and elsewhere. It should be said at once, lest the following criticisms obscure the appreciation, that the hopes thus raised are not disappointed. *Demand Analysis* is a highly instructive and provocative work that no economist or statistician could consult without profit, and for specialists it is indispensable.

In common with other branches of econometrics the study of consumers' behavior requires knowledge of the relevant chapters of economic theory and statistics in addition to skill in the interpretation and utilization of observational data. The book under review is organized around these three elements. After a first part which surveys the subject and summarizes the results there follow sections on the theory of choice, on stochastic processes, and on regression analysis: finally, some empirical investigations of the demand for foodstuffs in Sweden are discussed. Three of the five parts are completed by numerous exercises, some of which contain interesting new results.

ECONOMIC THEORY

The purpose of theory in an "applied" subject consists mainly in the formulation of (a) concepts in terms of which observations can be usefully described and (b) theorems, phrased in those terms, that allow statements about situations for which adequate observations are lacking. The usefulness of these concepts depends on their maintaining a certain invariance between different situations; thus, if the quantities bought by consumers showed little or no correlation with prices and incomes the concept of a "demand function" would not be a useful one.

* A review article of *Demand Analysis: A Study in Econometrics*, by Herman Wold in association with Lars Jureén. New York: John Wiley and Sons, Stockholm: Almquist and Wiksel, 1953. Pp. xvi, 358; $7.00.

Professor Wold is therefore rightly concerned to show from empirical evidence that demand functions, which are the cornerstone of consumption theory, do possess such stability. To what extent he has in fact shown this is a question to which we shall have to return.

The formulation of demand functions is not the ultimate aim of what the author describes as the "Paretoan" theory of consumer demand, though it was as far as the "demand function approach" (G. Cassel [3]) was prepared to go. In order to state theorems about these functions additional assumptions have to be made. These assumptions, which have been expressed in different ways, link the demand functions for an individual consumer with his preferences for various collections of goods. The now classical approach, due to Pareto and also favored by Wold, attributes to the consumer a consistent preference ordering for all such collections. A more recent version, advanced by Allen [1], assumes that the consumer only compares collections that are very close to each other;[1] this, however, is not a genuine generalization, for as soon as there is a finite difference between compared collections a chain of comparisons can be made and we are back to the preference ordering approach. A third approach, originally proposed by Samuelson [10], expresses consistency of preferences directly as a property of the demand functions; the reviewer has shown [6] that this "revealed preference" approach, when appropriately formulated, is also equivalent to the classical approach.

It has appeared worth while to go through these theoretical points because Professor Wold's discussion may easily mislead the unsuspecting reader. His theorem 4.6.1, in fact, seems to assert not only the equivalence of the marginal substitution and preference ordering approaches, but also of the latter and Cassel's demand function approach. In other words, the author does not regard the assumption that demand functions are derived from consistent preferences as an *additional* one; in his view these functions, lest they be "self-contradictory," must always satisfy the so-called "integrability condition," which expresses this consistency. Similarly, he declares the revealed preference approach to be merely a variation of the demand function approach. In doing so he misinterprets both, for it has been shown [6] that the strong axiom of revealed preference, which is certainly not satisfied by an arbitrary set of demand functions, is a necessary and sufficient condition for the existence of a consistent preference ordering. Using words

[1] Wold ascribes this "marginal substitution" approach also to Hicks, but this is very questionable. Even as early as the well-known pair of papers by Hicks and Allen of 1934 [5] a discrepancy between the views of the two authors was evident.

in their usual meanings there is nothing "self-contradictory" in a set of demand functions for which the integrability condition does not hold; all one can say is that the notion of preference does not apply to such a set. Wold's theorem 4.6.1 is therefore based on a *petitio principii.* If it holds true for the marginal substitution approach, this is only for the reasons given in the previous paragraph, and not because of Wold's circular argument.

In any review more space is inevitably devoted to criticism than to commendation, and we add at once that apart from this slip[3] the author's exposition of the preference ordering approach in Chapter 4 is lucid and original. Chapters 5, 6, and 7, dealing with the specification of demand patterns, relations between demand elasticities and market demand, lift consumption theory above the formal level on which it is too often discussed. Still more stress might have been laid, however, on the preponderance of corner equilibria and the resulting restrictions on the validity of traditional calculus methods. Curiously enough the author has failed to see that Hicks' method of deriving market demand is exactly the same as his own, so that his criticisms are unfounded. The exercise 2.27, which is claimed to illustrate Wold's objection, is highly instructive nevertheless. In Chapter 8 applications of preference theory to the supply of labor, to barter and to price index numbers are discussed.

Although Professor Wold points out and evaluates some of the limitations to Paretoan demand theory resulting from its static, non-stochastic and individualistic character, there is very little discussion of a possibly more serious complication, viz. that arising from consumers' assets. These assets, particularly in the form of durable consumption goods, lead to indivisibility problems and to the explicit introduction of time into the budget (see the recent work of Theil [13] and also Boulding [2]). Formally these problems may be covered by an extension of Wold's axioms, but in practice this is not very helpful; in fact the most difficult and interesting questions of theoretical and empirical demand research are precisely in this area. In a book entitled *Demand Analysis* readers should at least have been made aware of this field and referred, for instance, to the investigations of De Wolff [4] and Roos and Von Szeliski [9] on automobile demand. The preoccupation of the empirical chapters with food demand, though otherwise understandable, may also leave students with exaggerated notions as to the scope of a purely static approach.

---

[3] It might have been avoided if the author had paid more attention to Samuelson's questioning [11] of a similar theorem in [15]. As will be seen from the above Wold is also incorrect in describing the result in [6] as mathematically equivalent to his own assertion on the demand function approach.

STATISTICAL METHODS

The statistical controversies in demand analysis are perhaps of more importance than the theoretical questions reviewed above, since the former have a greater bearing on the numerical results obtained. Early in the history of econometrics it was recognized that a statistical apparatus designed mainly for biological experiments may not be entirely reliable in the analysis of economic observations. Two problems in particular have given rise to an extensive literature: serial interdependence in time-series and estimation in a system of simultaneous relations. On both of these topics, as well as on some related ones, Professor Wold has much of interest to contribute.

His discussion of time-series problems is based on a condensed but brilliant exposition of the theory of stationary processes, including a description of recent work by P. Whittle. He uses these results to show that in certain important cases classical least-squares methods retain their optimal properties in large samples, though the traditionally calculated standard errors may no longer indicate the goodness of fit. The special case to which the author devotes most attention is that of a recursive system. This is a system of behavior equations which can be solved successively rather than simultaneously. An example is the "pig cycle" model

$$d_t = D(p_t) \qquad s_t = S(p_{t-1}) \qquad p_t = p_{t-1} + \gamma(d_{t-1} - s_{t-1})$$

where d_t is demand, s_t supply, and p_t price at time t.

Professor Wold maintains—or at any rate leaves his reader with the impression—that such systems can be almost universally applied in economic analysis. It is not easy to summarize his argument, which is scattered over several chapters and appendix notes and sometimes hedged by qualifications that are not mentioned at a later stage. However, there can be no doubt that Wold's emphasis on recursive systems is an essential element in his sceptical attitude towards the simultaneous equations approach developed by Haavelmo, Koopmans, and others ([7], [8]). As is well known, the difficulties which prompted the latter development do not exist in recursive systems; more particularly, the coefficients in all their equations are identified and their least-squares estimates are asymptotically unbiased.

There are clearly desirable properties, and it is therefore necessary to consider how wide the scope of recursive systems in fact is. Strictly speaking they are indeed universal: there is every reason to believe that the economy, to use an anthropomorphic simile, solves its simultaneous equations by trial and error. Individuals adjust their actions to

parameters which they regard as fixed, but which are subsequently themselves affected by these actions so that new adjustments are required. Static models in which equilibrium values determine each other without sequence or lags are an abstraction, and have been recognized as such for almost as long as they have been used. This does not mean that they are useless, nor even that they are less realistic than recursive systems, in which such lags occur explicitly.

The problem here is that these lags are of very different length, ranging from a few seconds for the response of share or commodity prices to shifts in excess demand all the way to several years for the production of new ships or roads. In recursive systems of the kind described by Wold, however, lags have to be integral multiples of some unit period, usually the period to which the observations refer. If, as is usual in econometrics, annual observations have to be used, a one-year or two-year lag can easily be fitted in, but the question of what exactly should be done with lags of other lengths has not received Professor Wold's attention. In the absence of such a discussion his stress on recursive systems does not carry complete conviction.

We may perhaps detect here, as in other places, the results of an insufficient analysis of the use of approximative theoretical models in empirical research. The logic of regression analysis is treated with its application to experimental data as a prototype; with the aid of this interpretation rules for the selection of dependent variables are given. There is much play with the notion of causality, even though it has long lost its former pre-eminence in the physical and biological sciences and has never been very popular with economists, who tend to think in terms of functional rather than causal relationships.[3] Its introduction in any case hardly helps to bring out the difficulties peculiar to inference from non-experimental observations, arising mainly from the fact that the latter usually have to be taken as they come and are available in limited number only. Models therefore have to be chosen with reference to the data with which they are to be used. The resulting problem of how to choose between models is nowhere faced squarely in Wold's work, though there is a somewhat inconclusive discussion of the effects of additional regressors on the estimates of parameters already taken into account.

What is lacking, to put it in other words, is an adequate treatment of small-sample estimation. This compaint is of course not addressed to Professor Wold alone, for after Student's and Fisher's classical con-

---

[3] Recently an attempt to rehabilitate the notion of causality has been made by Simon [8], whose point of view is very similar to that of Wold.

tributions progress in this important area has been disappointing. Contemporary interest in estimation problems seems to be mainly centered on estimators with various asymptotic properties whose practical usefulness is often hard to see. Under these circumstances the author is unfortunately right in stating that for small samples we have to be satisfied with "the rough inference drawn by the use of large-sample methods," but this should not imply an abandonment of the search for more appropriate procedures.

These comments are mostly elicited by Wold's stimulating Chapter 2, in which he attempts to show that least-squares regression, despite the objections from Oslo and Chicago, is "essentially sound." The words in quotes are in fact characteristic of his attitude of militant conservatism on many points of controversy in statistics and elsewhere. Not all readers will find their doubts quieted by the forceful but occasionally one-sided array of arguments, but they will learn a great deal from trying to refute them.

EMPIRICAL FINDINGS

The empirical part, for which Mr. Jureén was jointly responsible, describes an extensive investigation of food demand in Sweden undertaken in connection with an inquiry into the long-term position of Swedish agriculture. Both family budgets and market statistics are used as source material. In line with Professor Wold's views on statistical methods as discussed above only a few technological innovations are to be noted. The numerical results, subject to the validity of the techniques employed, are on the whole very reasonable and their discussion, again with this proviso, is competent and illuminating. Our qualification refers to the authors' neglect of the supply side when dealing with demand equations, but this neglect is of course deliberate.

The work on family budgets (Chapter 16) is based on three Swedish surveys dating from 1913, 1923, and 1933. It is shown that estimates of national food consumption obtained by blowing up averages from the latter survey agree fairly closely with independent estimates of market demand. This is the more remarkable because the sample was not random but voluntary and participants had to keep detailed accounts for a whole year. It is in fact by no means clear that the voluntary approach is really inferior to the random sampling methods (with inevitably low response rates) currently in vogue.

From these budgets income elasticities are estimated for a considerable number of commodities and family types, quantity and expenditure elasticities being distinguished. In most of the analyses constant-

elasticity formulas are applied, but since the results reveal that the elasticities are not independent of income, some use is also made of the group of formulas suggested by Törnqvist. No standard errors of the estimates are calculated, on the ground that they are not theoretically justified for this material.

Earlier in the book (Chapter 14) there is a short discussion of equivalent adult scales where a method for determining the weights is proposed. This consists in calculating income elasticities in two ways: by pooling the separate estimates for different family types and by deriving a joint estimate for all households after their expenditures have been divided by the relevant number of equivalent adults. If the scale is correct, the two calculations will yield the same result. This agreement, however, is in general only a necessary and not a sufficient condition for correctness. Except in the case where only two kinds of persons (children and adults, for instance) are taken into account no unique scale will be obtained by Wold's method, and in practice it is desirable to specify many more categories of persons. Moreover, although Wold recognizes that the scale for total expenditures should be different from the scales for particular items this point does not seem to be allowed for in his method of computation.

Professor Wold's views on the relation between the income elasticities estimated from family budgets and those estimated from time series are also worth noting, especially since he was (in [14]) the first to combine the two sources in the manner now widely adopted. He distinguishes between short term and long term elasticities and maintains that the two sources both estimate the latter variety, which is usually the more interesting one. Because of the continuous introduction of new commodities, however, he thinks that the elasticities obtained from budget data on the whole tend to be smaller than those that refer to market statistics. This is an interesting observation, but it is not so clear that the time-series elasticity is really a long term figure and this somewhat weakens the author's conjecture.

In their work on market statistics (Chapter 17) Messrs. Jureén and Wold frequently use "conditional" regression analysis with income elasticities inserted as if they were known from other sources; they do not always use the estimates obtained from budget data but supplement these by "common sense" arguments. Sometimes they fix the price elasticity instead of the income elasticity. Standard errors of the estimates are calculated by means of a new formula which allows for autocorrelated disturbances. The symmetry of cross-price elasticities according to Slutsky and Hotelling is tested; it is found to hold rather

strikingly in the case of pork and beef, but much less so for animal and vegetable foodstuffs.

Some more general points in market demand are discussed in Chapter 15. Wold there pronounces himself against trend removal but in favor of deflating prices and incomes by a general price index. With the first advice the reviewer agrees, and, at any rate in the case of foodstuffs, also with the second. The author's argument on the latter question, however, is very superficial. He recommends deflation to correct for "changes in the monetary unit," and refers to Schultz [12] in support. Schultz, on the contrary, advocated deflation because it increases the degrees of freedom by one even though there is no exact method of taking changes in other prices into account. This is a much sounder argument, which shows incidentally that deflation is not invariably appropriate whereas Wold implies that it should always be applied.

In the final Chapter 18 the most interesting contribution is a detailed forecast of 1949–50 food consumption on the basis of pre-war demand functions tested against actual consumption in the forecast period. It is shown that the forecasts are on the whole reasonably accurate, and that in most cases they are nearer to the observations than the pre-war average on which a "naive" forecast might be based. It would be captious to deny that this satisfactory result speaks well for the validity of the methods used, whatever doubts one may have about their theoretical justification.

CONCLUSION

According to the preface *Demand Analysis* "is written in the dual form of a research report and a specialized textbook of econometrics." There are advantages and dangers in such a combination, particularly for the textbook half, and both are conspicuous here. The main advantage is that the methods discussed can be illustrated by actual applications though this is perhaps more effective if these applications belong to several fields instead of to a single rather narrow one as is the case here. Moreover readers will look in vain for actual applications of most of the economic theory and much of the statistics discussed in the earlier parts of the book. The dangers of the dual form are even more apparent. It has been pointed out already that the preoccupation with food demand in the empirical sections has led to an unfortunate neglect of dynamic factors. If industrial commodities had been studied as well as agricultural ones, there might also have been less inclination to ignore the complications due to simultaneous equations.

As a textbook *Demand Analysis* therefore has serious limitations,

which means that its use as such requires a considerable amount of additional explanation and amendment. The arrangement of the material could also be improved and much repetition eliminated. Professor Wold could hardly be blamed for not writing a standard work, since the subject is still too young to admit of one, but he would have come closer to writing one if he had been as successful in interpreting the ideas of others as he is in expounding his own. What we have here is essentially an admirable statement of the opinions and methods favored by one expert for his own research interests, and as such the book is an occasion for unqualified gratitude.

REFERENCES

[1] Allen, R. G. D., *Mathematical Analysis for Economists.* London, 1938.

[2] Boulding, K. E., *A Reconstruction of Economics.* New York, 1950.

[3] Cassel, G., *Theoretische Sozialökonomie.* Leipzig, 1918.

[4] De Wolff, P., "The demand for passenger cars in the U. S.," *Econometrica,* Vol. 6 (1938), 113–29.

[5] Hicks, J. R., and Allen, R. G. D., "A reconsideration of the theory of value," *Economica* N.S. 1 (1934), 52–75, 196–219.

[6] Houthakker, H. S., "Revealed preference and the utility function," *Economica* N.S. 17 (1950), 159–74.

[7] Koopmans, T. C. (ed.), *Statistical Inference in Dynamic Economic Models.* New York, 1950.

[8] Koopmans, T. C., and Hood, W. C., *Studies in Econometric Method.* New York, 1953.

[9] Roos, C. F., and Von Szeliski, V., *The Dynamics of Automobile Demand.* New York, 1939.

[10] Samuelson, P. A., "Note on the pure theory of consumer's behavior," *Economica,* N.S. 5 (1938), 51–71.

[11] Samuelson, P. A., "The integrability problem in utility theory," *Economica,* N.S. 17 (1950), 355–85.

[12] Schultz, H., *The Theory and Measurement of Demand,* Chicago, 1938.

[13] Theil, H., *The Influence of Stocks on Consumer's Behavior* (in Dutch). Amsterdam, 1951.

[14] Wold, H., *The Demand for Agricultural Products and Its Sensitivity to Price and Income Changes* (in Swedish). Stockholm, 1940.

[15] Wold, H., "A synthesis of pure demand analysis," *Skandinavisk Aktuarietidskrift,* 26 (1943), 85–118, 220–63 and 27 (1944), 69–120.

SOME PRACTICAL TECHNIQUES IN SERIAL NUMBER ANALYSIS

LEO A. GOODMAN
University of Chicago

The problem discussed is that of sampling from continuous and discrete uniform distributions. An application of this problem is presented which deals with the analysis of serial numbers on manufactured items in order to estimate the total number of items manufactured. Estimates of bounded relative error are obtained. Some justification for the use of these estimates is presented from the loss (cost) function point of view. Confidence intervals for the parameters are obtained and graphs are presented which may be used to determine the sample size required for confidence intervals of a given expected relative length. Tests of hypotheses are discussed. A method is presented for determining whether the serial numbers obtained are a random sample from a population of consecutive serial numbers.[1]

CONTENTS

[1] The author is indebted to Mr. William A. Aronow, New Holland Machine Company, and to Professor Harry V. Roberts, University of Chicgo, for very helpful comments.

1. INTRODUCTION

The analysis of serial numbers has several practical applications. We shall describe two such uses. The interested reader will no no doubt think of still other applications.

a) A commercial company could use the methods of serial number analysis in order to estimate the production and capacity of its compettitors. Representatives from the company could obtain the serial numbers of showroom equipment as well as equipment in use which has been produced by the competitors. Many of the basic methods have been developed for analyzing the serial numbers obtained by the company representatives (see [3]).

b) An organization has been using equipment which was purchased many years ago. The question was raised as to how many pieces of equipment had been purchased. No records were immediately available to determine the total purchase, since the purchase had been made years ago. Since serial numbers had been placed on each piece of equipment at the time of purchase, the serial numbers obtained from a sample of the equipment could be used to estimate the total purchase. Section 5 describes how this method was used to estimate the total number of pieces of equipment (desks, bookcases, etc.) which were purchased for the Division of the Social Sciences, The University of Chicago.

2. SUMMARY

Some of the practical problems which are of importance to organizations using "serial number analysis" will be considered here.

The arithmetic involved in the analysis of serial numbers seems to be simpler if the unknown production p is "assumed so large that variation is continuous" (see [3], p' 629). Some results for the "continuous variation" case will be presented which will serve as an approximation to the exact results. Some exact results will then be discussed.

The problem of obtaining confidence intervals for the total production p is studied. The sample size necessary to obtain confidence intervals of a given average relative length is determined. The power of tests of hypotheses concerning the true value of the production is also examined.

Rather than use an estimate of the production p which is unbiased or which minimizes the average of the squared error (see [3]) it might be desirable to have an estimate of which we are "almost certain" that it will be no more than, say, 1.2 times p and no less than, say, $0.8p$. The estimate which maximizes the probability of being included in the

desired interval may be determined. For example, if d is the difference between the largest and smallest serial number in a sample of thirty-one serial numbers, then we can be "99.99% confident" that the estimate $1.20d$ will be between $0.8p$ and $1.2p$. In other words, we can be "99.99% confident" that the relative error of the estimate $1.20d$ of p is less than .2. Justification of the use of such estimates of "bounded relative error" is presented within the framework of the theory of statistical decisions.

A method is also presented for testing the basic assumptions made in serial number analysis by examining the serial numbers which have been obtained. It is possible to test the hypothesis that the serial numbers obtained are a random sample. This method may also be used to detect whether there is a change in the procedure of serial numbering.

An application of the methods described herein is discussed in the final section.

3. CONTINUOUS VARIATION

In this section we shall assume that the serial numbers have a continuous uniform distribution between the initial serial number s and the final serial number $s+p$, where the total production p is unknown. Both the case when the initial serial number s is known and also when it is unknown will be considered.

3.1. *Initial Number Known*

When the initial number s is known, we might subtract s from each serial number obtained. The serial numbers (after the subtraction has been made) will then be uniformly distributed between 0 and p. The production p will be estimated using a sample of n serial numbers.

3.1.1. *Confidence intervals.* Let us first consider the problem of obtaining confidence intervals for p. If g is the largest serial number observed, suppose we state that "the total production p is between g and ag," where a is some constant greater than 1. Then the probability that this statement will be incorrect is $1/a^n$. That is, such a statement will be incorrect if and only if $ag<p(g<p/a)$. If $n=1$, the probability that $g<p/a$ is $\int_0^{p/a}dx/p=1/a$. Since each observation is independent, the probability that *all* observations, and therefore g in particular will be less than p/a is $1/a^n$. This probability $1/a^n=\alpha$ of making an incorrect statement may be made small by choosing a large value for the constant a, or by obtaining a large sample of n serial numbers. We might first determine how small the probability α of making an incorrect statement should be, and then determine a or n from the relation

$\alpha = 1/a^n$. The interval "g to ag" in which it is stated that p lies is called the "$(1-\alpha)\cdot 100\%$ confidence interval" since the probability is $1-\alpha$ that the statement will be correct.

The length of the confidence interval in which it is stated that p lies is $ag - g = g(a-1)$. Since the expected value of g is $pn/(n+1)$, the expected length of the interval is $pn(a-1)/(n+1)$. The expected relative length of the interval is $n(a-1)/(n+1) = \lambda$. We might first determine how small the probability α of making an incorrect statement should be and also how small the expected relative length λ of the confidence interval should be. The sample size n of the serial numbers may then be determined by the relations

$$n(a-1)/(n+1) = \lambda \quad \text{and} \quad \alpha = 1/a^n \quad \text{or}$$

$$a = \lambda + 1 + \lambda x \quad \text{and} \quad a = \alpha^{-x}, \quad \text{where} \quad x = 1/n.$$

For any given values of α and λ, graphs of the functions $\lambda + 1 + \lambda x$ and α^{-x} can be drawn. The value x_0 of x where the two graphs intersect is then the desired solution of the last two equations. The reciprocal $1/x_0 = n_0$ of this solution is the desired sample size. If then n_0 serial numbers are obtained, we will have $(1-\alpha)\cdot 100\%$ confidence in the statement that "p lies between g and ag." The expected relative length of this confidence interval is the desired value λ.

It is interesting to note that among all $(1-\alpha)\cdot 100\%$ confidence intervals of the form "$a_1 g$ to $a_2 g$," where $1 \leqq a_1 < a_2$, the confidence interval with the smallest average length is obtained by taking $a_1 = 1$, which is what we have done.

3.1.2. *Testing hypotheses.* Let us now consider the problem of testing the hypothesis that the total production is a given value p_0. This hypothesis will be rejected when the given value p_0 does not lie within the confidence interval. In other words, having observed a sample of serial numbers, we make a confidence statement that "p is between g and ag," and reject the "null" hypothesis that the total production is a given value p_0 if this value lies outside the confidence interval. The probability is $\alpha = 1/a^n$ of rejecting this hypothesis when it is in fact true. We should like the probability of rejecting the null hypothesis (that the total production is p_0) to be large, when the hypothesis is in fact false (i.e., when the total production is a value p different from p_0). This probability $1-\beta$ of correctly rejecting the null hypothesis, when in fact the true production is p, may be determined by the following formula:

$$1-\beta(p) = \begin{cases} 1, & \text{when} \quad p < p_0/a \\ \alpha(p_0/p)^n, & \text{when} \quad p_0/a \leqq p \leqq p_0 \\ 1-(1-\alpha)(p_0/p)^n, & \text{when} \quad p > p_0. \end{cases}$$

We call $1-\beta(p)$ the power function of the test.

The formula for the power function $1-\beta(p)$ follows directly from the following considerations. The null hypothesis that the total production is a given value p_0 will be rejected whenever $p_0<g$ or $p_0>ag$. But $g<p_0/a$ if and only if all observations are less than p_0/a. The probability that an observation will be less than p_0/a is p_0/ap, when in fact the true production is $p>p_0/a$. Hence the probability that all observations will be less than p_0/a (i.e., $g<p_0/a$), is $(p_0/ap)^n=(p_0/p)^n\alpha$ if $p>p_0/a$. If $p<p_0/a$, rejection of the null hypothesis is certain since $g\leqq p<p_0/a$. The probability that at least one observation will be greater than p_0 (i.e., $g>p_0$) is zero for $p<p_0$, and it is $1-(p_0/p)^n$ for $p>p_0$. From these conclusions the formula for the power function follows directly.

We might first determine how small the probability α of incorrectly rejecting the null hypothesis should be and also how large the probability $1-\beta$ should be of correctly rejecting the null hypothesis when a particular alternative hypothesis $p=p_1$ (different from p_0) is true. If the alternate hypothesis $p=p_1$ has been specified the appropriate sample size of the serial numbers required can be determined by solving the equation

$$1-\beta = 1-\beta(p_1)$$

for the value of n. For example, if $p_0\alpha\leqq p_1\leqq p_0$, then

$$1-\beta = \alpha(p_0/p_1)^n$$

$$(1-\beta)/\alpha = (p_0/p_1)^n$$

or

$$n = \log\,[(1-\beta)/\alpha]/\log\,[p_0/p_1].$$

3.1.3. *Estimates of bounded relative error.* In [3], the problem of point estimation of p was considered and the unbiased estimate of p which had the smallest variance was given. The relation between this unbiased estimate and various other point estimates of p was examined. The problem of point estimation will now be considered from a somewhat different point of view. We might want to be "almost certain" that the estimate of production p obtained from the sample of n serial numbers will not be more than 1.2 times as large as the true production p, and

will not be smaller than $0.8p$. If the estimate is of the form cg, where $c \geqq 1$ is a constant and g is the largest among the n serial numbers, then the probability that the estimate cg will lie between $0.8p$ and $1.2p$ is

$$(1.2/c)^n - (0.8/c)^n, \quad \text{when} \quad c \geqq 1.2$$

and

$$1 - (0.8/c)^n, \quad \text{when} \quad c \leqq 1.2.$$

Hence the probability that cg will lie between $0.8p$ and $1.2p$ is maximized when $c = 1.2$ and, in that case, the probability is

$$1 - (0.8/1.2)^n.$$

The sample size n necessary in order that we can be "$(1-\alpha)\cdot 100\%$ confident" that $1.2g$ lies between $0.8p$ and $1.2p$ is determined by the relation

$$1 - \alpha = 1 - (0.8/1.2)^n$$

or

$$n = \log \alpha / \log (0.8/1.2).$$

It may be desirable to determine an interval c_1g to $c_2g (1 \leqq c_1 \leqq c_2)$ of which we can be at least "$(1-\alpha)\cdot 100\%$ confident" that any given estimate of the form cg in that interval $(c_1 \leqq c \leqq c_2)$ will lie between $0.8p$ and $1.2p$. In order to obtain such an interval, it is clear that the sample size n must be greater than $\log \alpha / \log (0.8/1.2)$. In that case, the values of c_1 and c_2 are determined by

$$1 - \alpha = 1 - (0.8/c_1)^n$$

and

$$1 - \alpha = [(1.2)^n - (0.8)^n]/c_2^n, \quad \text{since} \quad c_1 \leqq c \leqq c_2.$$

We might wish to determine an interval c_3g to c_4g of which we can be "$(1-\alpha)\cdot 100\%$ confident" that the entire interval will lie between $0.8p$ and $1.2p$. If $n > \log \alpha / \log (0.8/1.2)$, appropriate values of $c_3 < 1.2$ and $c_4 > 1.2$ can be determined by the relation

$$(1.2/c_4)^n - (0.8/c_3)^n = 1 - \alpha.$$

More generally, if an estimate cg is desired which maximizes the probability of being included between k_1p and k_2p (where the k's are

given constants such that $k_1 < k_2$) then the estimate should be $k_2 p$. If the sample size n is greater than log α/log (k_1/k_2), then the probability is at least $1-\alpha$ that any given estimate of the form g times a given constant in the interval $c_1 g$ and $c_2 g$ will lie between $k_1 p$ and $k_2 p$, where

$$c_1{}^n = k_1{}^n/\alpha$$

and

$$c_2{}^n = [k_2{}^n - k_1{}^n]/(1-\alpha).$$

Also, the probability is $1-\alpha$ that the entire interval $c_3 g$ to $c_4 g$ will lie between $k_1 p$ and $k_2 p$ where

$$(k_2/c_4)^n - (k_1/c_3)^n = 1 - \alpha.$$

In practice it may sometimes be possible to determine the constants k_1 and k_2 so that if the estimate $\hat{p}$ of p is between $k_1 p$ and $k_2 p$ it will be "close enough." By "close enough" we mean that no loss is incurred when an estimate $\hat{p}$ of p is made which is between $k_1 p$ and $k_2 p$. When the estimate $\hat{p}$ is not between $k_1 p$ and $k_2 p$, then the loss incurred in using an estimate which is not "close enough" may be some given constant, say, 1. If the loss incurred in estimating p by $\hat{p}$ may in fact be described by the function

$$L(\hat{p}, p) = \begin{cases} 0 & \text{when } \quad k_1 p < \hat{p} < k_2 p, \\ 1 & \text{otherwise,} \end{cases}$$

then the estimate which maximizes the chance of being included between $k_1 p$ and $k_2 p$ also minimizes the expected loss. Hence the estimate $k_2 g$ which maximizes the chance of being included between $k_1 p$ and $k_2 p$ may be justified within the framework of the theory of statistical decisions. For a more general discussion of the problem treated in this paragraph the reader is referred to [2].

3.1.4. *Tests of randomness and consecutive serial numbering.* It has been assumed herein that the n serial numbers obtained are a random sample from all the serial numbers which are distributed uniformly (numbered consecutively) between the initial serial number s and the final serial number $s+p$, where s or $s+p$ (or both) may be unknown. Before applying the statistical methods which have been based on this assumption, it is desirable to examine the sample of n serial numbers and test whether this assumption is justified. That is, the hypothesis that the serial numbers were obtained from a random sample of n observations from a uniform distribution between s and $s+p$ should

be tested. The question "Are the serial numbers a random sample?" will be studied.

When the initial serial number s is known, it has been assumed that the serial numbers (after s has been subtracted from each serial number) are uniformly distributed between 0 and p, where p is unknown. The n serial numbers have been assumed to be a random sample of serial numbers. Let us now consider the problem of testing the hypothesis that the n serial numbers are a random sample. We note that the hypothesis to be tested is not concerned with determining the unknown true value of the production p. Several tests are available for the hypothesis that the n serial numbers are a random sample from all the serial numbers uniformly distributed between 0 and p, where p is not specified. Consider all n serial numbers obtained except the largest serial number g. If the hypothesis to be tested is true, then this sample of the $n-1$ smallest serial numbers will be uniformly distributed between 0 and g, when g is given. Hence, dividing these $n-1$ serial numbers by g, the numbers obtained will be uniformly distributed between 0 and 1, when the hypothesis to be tested is true. In order to test the hypothesis of randomness, we might test whether these $n-1$ serial numbers (divided by g) are uniformly distributed between 0 and 1. This can be done using the Kolmogorov statistic or one of the other statistics (e.g., chi-square, maximum difference, etc.) described in [1]. For example, if $n=31$, a graph of the sample cumulative distribution of the $n-1=30$ smallest serial numbers obtained (when divided by the largest serial number obtained can be drawn). The maximum absolute difference between this sample cumulative and the cumulative of the uniform distribution (the diagonal line) is then determined. From Table 1 ($N=30$), on page 428 of [1], we find that the probability is .97745 that this maximum absolute difference between the cumulatives will be less than 8/30. Hence, if a test is to be performed at the .02255 level of significance, we will accept the hypothesis of randomness whenever the maximum absolute difference between the cumulatives is less than 8/30.

If the hypothesis of randomness is accepted, the analysis described in the preceding sections herein and in [3] could then be used. If the hypothesis is rejected, the sample of serial numbers should be examined to determine what is nonrandom about it. On the basis of such an inquiry *ad hoc* methods for estimating the true production p could be determined.

This approach may also be used to see whether there are changes in

the procedure of serial numbering. If the procedure changes (i.e., if the serial numbers are not uniformly distributed between the initial serial number and the final serial number), then a random sample of the serial numbers might indicate a nonuniform distribution. The test proposed in this section may be considered as a test of the hypothesis that serial numbering was done consecutively, as well as a "test of randomness."

3.2. *Initial Number Unknown*

3.2.1. *Confidence intervals.* Let us first consider the problem of obtaining confidence intervals for p.

The probability that the difference d between the largest and smallest among the n serial numbers is greater than $p/b(b \geqq 1)$ may be determined by the following relation (see [4], page 386):

$$\begin{aligned} \Pr\{p \geqq d \geqq p/b\} &= \Pr\{1 \geqq d/p \geqq 1/b\} \\ &= \int_{1/b}^{1} n(n-1)z^{n-2}(1-z)dz \\ &= 1 - nb^{1-n} + (n-1)b^{-n} \\ &= \Pr\{d \leqq p \leqq bd\}. \end{aligned}$$

Suppose the statement is made that "the total production p is between d and bd," where b is some constant. Then the probability α that this statement will be incorrect is $nb^{1-n}+(1-n)b^{-n}=\alpha$. This probability α of making an incorrect statement may be made small by choosing a large value for the constant b, or by obtaining a large sample of n serial numbers. We might first determine how small the probability α of making an incorrect statement should be, and then determine b or n from the relation $\alpha=nb^{1-n}+(1-n)b^{-n}$. Tables are available which will simplify the computations (see [5], [6]). A reprint of [6] may be purchased from *Biometrika*.

Let us illustrate the methods just described by a numerical example. If α is chosen equal to 0.05, the value of $1/b$ can be determined from the entries in column $4=v_1$ on p. 174 of [6] where $2(n-1)=v_2$. If $n=31$ serial numbers have been obtained, then $1/b$ is determined by the entry in the fourth column ($v_1=4$) and third row from the bottom ($v_2=60$) of the table on page 174 in [6]. Hence $1/b=.85591$ and $b=1.17$. Upon observing 31 serial numbers, we will be 95% confident in the statement that "the total production p lies between d and $1.17d$."

The length of the 95% confidence interval for $n=31$ serial numbers is $d(1.17-1)=0.17d$. Since the expected value of d is $p(n-1)/(n+1)$,[2] the expected length of the interval is $(0.17)\ p(n-1)/(n+1)=0.16p$. The expected relative length of the interval is $\lambda=0.16$. We might first determine how small the probability α of making an incorrect statement should be and also how small the expected relative length λ of the confidence interval should be. Then the relations

$$(b-1)(n-1)/(n+1)=\lambda \quad \text{and} \quad nb^{1-n}+(1-n)b^{-n}=\alpha$$

can be used to determine b and the necessary sample size n. Writing $1/b=y$ and $1/(n-1)=x$, the first relation can be replaced by $1/y=\lambda+1+2\lambda x$.

Other methods for determining n may also be used; e.g., successive approximation procedures.

3.2.2. *Testing hypotheses.* The problem of testing the hypothesis that the total production is a given value p_0 may be studied in the same way as was done in Section 3.1.2. Direct computations may be made for any test at a given level α of significance in order to determine the power function of the test. The tables in [5] and [6] may be used to simplify computation.

3.2.3. *Estimates of bounded relative error.* Let us now consider the problem of point estimation of p from the same point of view as in Section 3.1.3. We might want to be "almost certain" that the estimate of p obtained from the sample of n serial numbers "will not be more than $1.2p$ nor smaller than $0.8p$." If the estimate is of the form cd, where $c>1$ is a constant and d is the difference between the largest and smallest among the n observed serial numbers, then the probability that the estimate will be between $0.8p$ and $1.2p$ is maximized when

$$(1.2)^{n-1}(1-1.2/c)-(0.8)^{n-1}(1-0.8/c)=0$$

or when

$$c=[(1.2)^n-(0.8)^n]/[(1.2)^{n-1}-(0.8)^{n-1}]. \tag{1}$$

The sample size n necessary in order that we can be "$(1-\alpha)\cdot 100\%$ confident," that cg lies between $0.8p$ and $1.2p$ is determined by the relation

[2] The reader will notice that the expected value of d presented on page 627 of [3] is $(p+1)(n-1)/(n+1)$. The formula in [3] was derived for the exact model whereas the formula in this text is for the continuous variation model. Hence, $d(n+1)/(n-1)-1$ is the unbiased estimate of p in the exact model (see [3]) whereas the unbiased estimate of p for the continuous variation model is $d(n+1)/(n-1)$.

$$1-\alpha = \int_{0.8/c}^{1.2/c} n(n-1)z^{n-2}(1-z)dz$$
(2)
$$= \Pr\{0.8/c \leqq d/p \leqq 1.2/c\}$$

and relation (1).

If the sample size is larger than the sample required by the preceding relations (1) and (2), two constants $c_1 \leqq c$ and $c_2 > c$ can be determined, where c is defined by relation (1), such that we can be at least "$(1-\alpha)\cdot 100\%$ confident" that any given estimate of the form d times a given constant in the interval $c_1 d$ to $c_2 d$ will lie between $0.8p$ and $1.2p$. The values of c_1 and c_2 are determined by the relations

$$1-\alpha = \Pr\{0.8/c_2 \leqq d/p \leqq 1.2/c_2\}$$

and

$$1-\alpha = \Pr\{0.8/c_1 \leqq d/p\}.$$

It may be desirable to determine an interval $c_3 d$ to $c_4 d$ of which we can be "$(1-\alpha)\cdot 100\%$ confident" that the entire interval will lie between $0.8p$ and $1.2p$. When the sample size n is larger than the sample required by relations (1) and (2), appropriate values of $c_3 < c$ and $c_4 > c$ may be determined by the relation

$$1-\alpha = \Pr\{0.8/c_3 \leqq d/p \leqq 1.2/c_4\}.$$

The numbers 0.8 and 1.2 can be replaced by k_1 and k_2 respectively in the preceding discussion to obtain more general results. A justification of estimates of bounded relative error may be presented, as was done in Section 3.1.3, within the framework of the theory of statistical decisions. The estimate cd which maximizes the chance of being included within k_1p and k_2p is also the estimate which minimizes the expected loss if no loss is incurred when the estimate is within k_1p and k_2p and a constant loss is incurred otherwise.

Let us illustrate the computations required in the preceding discussion by considering a sample of $n=31$ serial numbers. The value of c as defined by relation (1) is equal to 1.20 (to three significant digits), when $n=31$. Hence, the estimate $1.20\,d$ maximizes the chance of being included between $0.8p$ and $1.2p$. From the tables on page 54 of [5] we find that the chance is .9999 that $1.20d$ will lie between $0.8p$ and $1.2p$.

Suppose we wish to be 95% confident of all statements made, i.e., $\alpha=.05$. The second column ($p=30$) of the table ($q=2$) on page 54 of [5] presents the distribution of d. Using this information together with

the entry in the eighth column ($\nu_1=60$) and the fourth row ($\nu_2=4$) on page 175 of [6], we see that c_2 is about $1.2/(1-.011585)=1.21$ (to three significant digits). Hence if the estimate of production p based on 31 serial numbers is $1.21d$, then the probability is 0.95 that this estimate will be between $0.8p$ and $1.2p$. From the table on page 174 of [6] ($\nu_1=4$, $\nu_2=60$) we see that

$$\Pr\{0.8 \leqq d/p\} > .95.$$

Hence, we are at least 95% confident that any given estimate (of the form d times a given constant) in the interval d and $1.21d$ will lie between $0.8p$ and $1.2p$. We also find from the tables that the probability is about .95 that the entire interval d to $1.21d$ will lie between $0.8p$ and $1.2p$.

3.2.4. *Tests of randomness and consecutive serial numbering.* Let us consider the hypothesis that the n serial numbers obtained are a random sample from the population of uniformly distributed serial numbers. In the case where the initial number is unknown, we consider all n serial numbers obtained except the largest serial number g and the smallest serial number f. If the hypothesis to be tested is true, then this sample of $n-2$ serial numbers (all except g and f) will be uniformly distributed between f and g, when f and g are given. Hence, subtracting f from these $n-2$ serial numbers and then dividing the numbers obtained by $g-f$, the adjusted numbers will be uniformly distributed between 0 and 1, when the hypothesis to be tested is true. In order to test the hypothesis of randomness, we might test whether these $n-2$ adjusted serial numbers (when f is subtracted from the serial numbers and the numbers obtained are then divided by $g-f$) are uniformly distributed between 0 and 1. This can be done using the Kolmogorov statistic or one of the other statistics (e.g., chi-square, maximum difference, etc.) as mentioned in Section 3.14. For example if $n=31$, the sample cumulative distribution of the $n-2=29$ adjusted serial numbers obtained can be graphed. The maximum absolute difference between this sample cumulative and the cumulative of the uniform distribution (the diagonal line) can then be determined. From Table 1 ($N=29$) on page 428 of [1], we note that the probability is .98076 that this maximum absolute difference between the cumulatives will be less than 8/29. Hence, if a test is to be performed at the .01924 level of significance, the hypothesis of randomness and consecutive (uniformly distributed) serial numbers will be accepted whenever the maximum absolute difference between the cumulatives is less than 8/29.

4. THE EXACT MODEL

In the preceding sections we have assumed that the serial numbers have a continuous uniform distribution between the initial serial number s and the final serial number $s+p$. This was done in order to simplify the problem and because for practical problems (when the value of p is large) the results obtained will serve as an approximation to results for the exact model of a discrete, finite, uniform population (see [3]).

On page 624 of [3], the exact confidence intervals and tests of hypotheses are obtained for the case where the initial serial number is known. Since exact confidence intervals and tests of hypotheses were not discussed in [3] for the case where the initial serial number is unknown, we shall now consider that problem.

From [3], we see that the probability that the difference d between the largest and smallest among n serial numbers will be less than or equal to a given constant c may be determined from the relation

$$\Pr\{d \leqq c \mid n, p\} = \sum_{d=n-1}^{c} n^{(2)}(d-1)^{(n-2)}(p-d)/p^{(n)}$$

$$= nc^{(n-1)}/(p-1)^{(n-1)} - (n-1)(c+1)^{(n)}/p^{(n)},$$

where $c^{(m)} = c!/(c-m)!$. As a first approximation to this probability we might replace the exact model by the model of a continuous uniform distribution and obtain $Pr^1\{d \leqq c \mid n, p\} = n(c/p)^{n-1} - (n-1)(c/p)^n$ for which convenient tables are available (see [5] and [6]).

Suppose we wish to test the null hypothesis that $p = p_0$ against the alternative hypothesis $p > p_0$. Then the rejection region for a significance test at level α is obviously $d > c_1 + 1$ where c_1 is the largest integer satisfying

$$\Pr\{d \leqq c_1 \mid n, p_0\} < 1 - \alpha.$$

If we wish to test the null hypothesis $p = p_0$ against the alternative hypothesis $p \leqq p_0$, then the rejection region for a significance test at level α is $d \leqq c_2$ where c_2 is the smallest integer satisfying

$$\Pr\{d \leqq c_2 \mid n, p_0\} > \alpha.$$

A two-sided test at level α of the null hypothesis $p = p_0$ against the two-sided alternative $p \neq p_0$ is defined by the acceptance region $c_2 \leqq d \leqq p_0 - 1$. A two-sided test at the 2α level might be based on the acceptance region $c_2 \leqq d \leqq c_1 + 1$.

The results of the preceding paragraph may now be used to obtain confidence intervals. That is, the left-sided $1-\alpha$ confidence interval is $p \geqq k_1$, where k_1 is the smallest integer satisfying

$$\Pr\{d < d_0 | n, k_1\} < 1 - \alpha,$$

and d_0 is the actual difference between the largest and smallest among the n serial numbers observed. The right-sided $1-\alpha$ confidence interval is $p \leqq k_2$, where k_2 is the largest integer satisfying

$$\Pr\{d \leqq d_0 | n, k_2\} > \alpha.$$

A two-sided $1-\alpha$ confidence interval is $d+1 \leqq p \leqq k_2$ and a two-sided $1-2\alpha$ confidence interval is $k_1 \leqq p \leqq k_2$.

5. AN APPLICATION

The Division of the Social Sciences of the University of Chicago has been using equipment (desks, bookcases, etc.) upon which serial numbers had been placed. The question was raised as to how many such pieces of equipment were there.

The serial numbers on thirty-one pieces of equipment were observed. The 31 serial numbers obtained were:

83, 135, 274, 380, 668, 895, 955, 964, 1113, 1174, 1210, 1344, 1387, 1414, 1610, 1668, 1689, 1756, 1865, 1874, 1880, 1936, 2005, 2006, 2065, 2157, 2220, 2224, 2396, 2543, 2787.

The serial numbers range from 83 to 2787. The sample cumulative distribution of the 29 serial numbers obtained between the smallest and largest serial numbers is graphed in Figure 5.1. The diagonal line in Figure 5.1 represents the uniform cumulative distribution between the smallest serial number 83 and the largest serial number 2787. From Figure 5.1 we see that the maximum absolute difference between the two cumulative distributions is $(9.65-5)/29 = .16$. If the serial numbers obtained are a random sample from a population of uniformly distributed serial numbers, then there is more than a $1-.68280 = .3172$ probability of obtaining a maximum absolute difference of .16 or larger (see page 428, Table 1, $N=29$, in [1]). Hence the null hypothesis that the serial numbers obtained are a random sample from a population of consecutive serial numbers is accepted.

From Section 3.2.1 we see that the unbiased estimate of the total number p of pieces of equipment is $d\ 32/30 = (2787-83)32/30 = (2704)32/30 = 86528/30 = 2884.3$ for the continuous variation model (2883.3 for the exact model). Also, the 95% confidence interval for p is "$2704 \leqq p \leqq 1.17(2704)$" or "$2704 \leqq p \leqq 3163.7$."

From Section 3.2.3 we see that the chance is .9999 that the estimate

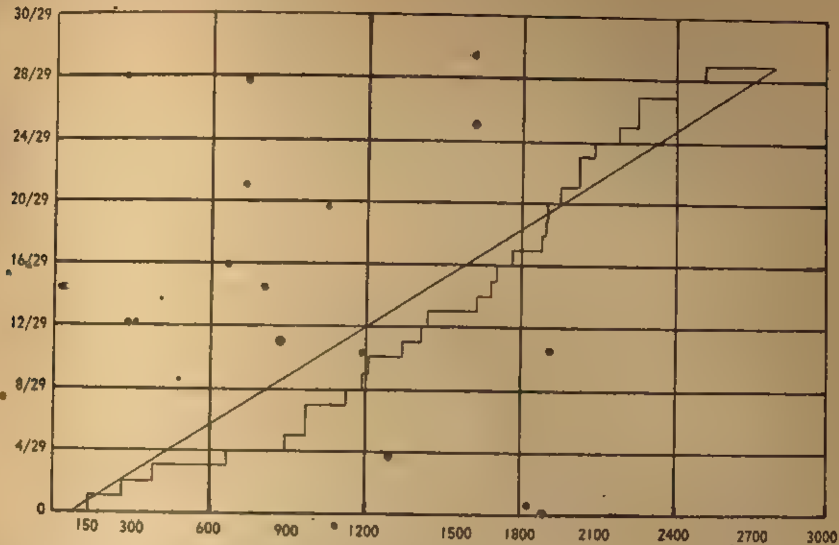


FIG. 5.1. Sample Cumulative Distribution of the 29 Observed Serial Numbers Between the Smallest and Largest.

$1.20d = 1.20(2704) = 3244.8$ will be within 20 per cent of p. This estimate minimizes the expected loss if no loss is incurred when the estimate is within 20 per cent of p and a constant loss is incurred otherwise. The probability is .95 that the estimate $1.21d = 1.21(2704) = 3271.8$ will be within 20 per cent of p. In fact the probability is appropriately .95 chance that the entire interval d to $1.21d$, or 2704 to 3271.8 will lie within 20 per cent of p.

It was a relatively simple task to obtain the serial numbers of 31 pieces of equipment and then to estimate p in the manner described herein. Determining the true value of p (the total number of pieces of equipment) was much more time consuming. These pieces of equipment had been purchased in the period between 1928 and 1934 and no records were immediately available to determine the total purchase. We are indebted to Mrs. Ruth Denney, Aministrative Assistant to the Dean of the Social Sciences. After several days and many inquiries, Mrs. Denny was able to locate the records and found that the total number p of pieces of equipment was *2885*.

6. REFERENCES

[1] Birnbaum, Z. W., "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size," *Journal of the American Statistical Association*, 47 (1952), 425–41.

[2] Blackwell, D., and M. A. Girshick, *Theory of Games and Statistical Decisions*, in press.

[3] Goodman, Leo A., "Serial number analysis," *Journal of the American Statistical Association*, 47 (1952), 622–34.
[4] Mood, A. M., *Introduction to the Theory of Statistics*, McGraw-Hill Book Company, New York, 1950.
[5] Pearson, Karl, *Tables of the Incomplete Beta Function*, Cambridge University Press, London, 1932.
[6] Thompson, Catherine M., "Tables of percentage points of the incomplete beta function," *Biometrika*, 32 (1941), 151–81.

THE PROBLEM OF AUTOCORRELATION IN REGRESSION ANALYSIS*

R. L. ANDERSON
North Carolina State College

1. INTRODUCTION

IN LEAST squares analysis, the usual regression model is

$$Y_t = \beta_0 + \sum_{i=1}^{r} \beta_i X_{it} + \epsilon_t, \qquad t = 1, 2, \cdots, n,$$

where the predictors, the X's, are assumed fixed in repeated sampling and the ϵ's independently distributed with the same variance, σ^2. The X's may be merely dummy variates (0 or 1), as in classification data (often called analysis of variance data). When tests of significance or confidence limits for the parameters are used, one usually assumes normality of the ϵ's. Even if the X's and Y follow a multivariate normal distribution, the least squares point and interval estimates of the β's can be used, and the usual null tests applied.

If the nY's are successive observations in time, the experimenter frequently wishes to investigate the nature of the response curve over time. In this case he might set $X_{it} = t^i$, or he might use the method of *harmonic analysis* to search for periodicities in Y. In other cases, the assumed model might involve lagged values of Y as predictors. For example,

$$Y_t = \beta_0 + \sum_{i=1}^{r} \beta_i Y_{t-i} + \epsilon_t. \tag{1}$$

This is an *autoregressive model*. Finally one could use a combined regression model with lagged Y's, present X's, lagged X's, and time as predictors. The method of least squares is applicable for autoregressive models, provided n is large [see Mann and Wald [6]].

One of the major difficulties with the use of least squares methods with time series is the strong possibility that the ϵ's are not independent. Aitken [1] pointed out that it is correlation of the ϵ's and not of the Y's which is to be avoided. It is possible that if the X's and Y's are both correlated in time, the errors will be relatively uncorrelated. A considerable amount of research has been devoted to the problem

* Paper presented at Annual Meeting of American Statistical Association, Chicago, December 27, 1952.

of testing for the existence of correlation in the errors, but all too little on the more important problem of the best estimation procedure when correlations do exist. Summaries of current methods of analyzing time series are given by Kendall [5] and Tintner [7].

The correlation of successive items in a time series was called a *lagged serial correlation* by Yule [9]. At the present time, it is more popular to use the term serial correlation to apply to the correlation between two series and the word *autocorrelation* for this correlation between successive items in a given series [see, for example, Tintner [7]]. I shall use this distinction. Many of the earlier papers on this subject, however, use the Yule terminology, as can be noted from the Bibliography. If we have a set of equally spaced values, $Z_1, Z_2, \cdots, Z_n$, selected from a population with zero mean, the autocorrelation coefficient of lag L is

$$r_L = \frac{SZ_iZ_{i+L}}{\sqrt{SZ_i^2SZ_{i+L}^2}}, \tag{2}$$

where i goes from 1 to $n-L$.[1] Most writers have preferred to use a definition in which the denominator is simply

$$\mathop{S}_{i=1}^{n} Z_i^2. \tag{3}$$

A symposium on autocorrelated time series analysis was held in 1946 under the auspices of the Royal Statistical Society. M. S. Bartlett [2] presented a general paper and Foster [4] and Cunningham and Hynd [3] presented papers on the use of autocorrelation methods in non-economic fields. J. W. Tukey at the 1951 Annual Meeting of the American Statistical Association proposed the use of the autocovariance (the numerator of r_L) in his method of spectrum analysis of time series.

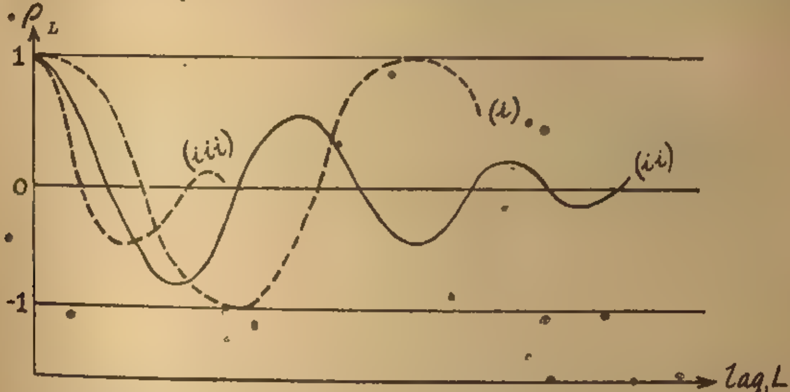
2. TESTS OF SIGNIFICANCE FOR AUTOCORRELATION

Yule [10] showed that the distribution of the correlation between two autocorrelated series tends to be U-shaped with a majority of the correlations near ± 1. Bartlett [15] said that if the errors were autocorrelated, we could use the usual tests of significance of regression coefficients on a preliminary basis. If these coefficients were non-significant, accept the result; if they were significant, a test was needed which took account of the autocorrelation.

[1] S will be used to indicate summation over sample values.

One of the common methods of analyzing a single time series is *harmonic analysis*, in which the X's are sine and cosine terms. Fisher [18] presented a test of significance of the various amplitudes (the β's), in the restricted case of independent errors. Wilson [41] suggested that one compute successive lagged autocorrelation coefficients until the first non-significant one is reached; then use this lag (L) as an indication of the proportion of independent observations ($1/L$).

Three possible models are used to explain stationary trend-free time series data. Wold [8] indicated that the choice depended upon the relationship of successive true autocorrelation coefficients, ρ_L. These are usually displayed in a *correlogram*, as shown in the figure below.



(i) Repeated non-damped cycles: use harmonic analysis.
(ii) Damped correlations but with $|\rho| > 0$: use linear autoregression.
(iii) Damped correlations, with $\rho_L = 0$ for $L > m$: use the method of moving averages,

$$Y_t = e_t + \sum_{s=1}^{m} \gamma_s e_{t-s}. \tag{4}$$

Tintner [7] also discusses these methods in detail. Bartlett [2] cautions about the use of empirical correlograms to determine the correct model because successive sample autocorrelation coefficients tend to be highly correlated.

The approximate test of amplitudes in harmonic analysis and the decision regarding a proper model depends upon a test of significance for autocorrelation. For this reason the author decided to work on the distribution of r_L in 1939. Because of the mathematical difficulties in-

volved, it was decided to follow up a suggestion of Hotelling to use a circular definition

$$r_L' = \frac{SZ_iZ_{i+L}}{SZ_i^2}, \tag{5}$$

where i goes from 1 to n, and $Z_{n+k}=Z_k$.

In 1941, the author studied the distribution for normal Z's when the population mean was zero, and, in 1942, the distribution for Z's which were deviations from the sample mean. Significance levels were computed for r_1', and for several cases of lags greater than one. The theory was simplified by the fact that

$$r' = \frac{\sum \lambda_i m_i}{\sum_i m_i},$$

where the m's are χ^2 variables with one degree of freedom, and the λ's are latent roots of the characteristic equation of the matrix of the coefficients in the numerator. Koopmans [22] reported on the distribution of r_1, as an estimate of ρ in the simple autoregressive model:

$$Y_t = \rho Y_{t-1} + \epsilon_t. \tag{6}$$

At the same time Dixon [16] was studying the moments of the distribution of r_1' and used Beta approximations to the exact distributions to obtain significance levels. T. W. Anderson [14] later showed that no test of the hypothesis $\rho=0$ exists which is uniformly most powerful against alternatives of the Koopmans type.

Sometime before this, a problem involving autocorrelation came up in industrial quality control, in which the mean tended to creep up and down slightly on successive observations. In order to study the variation in the production process, von Neumann, *et al.* [35] suggested that the statistic

$$\frac{\delta^2}{2} = \mathop{S}_{i=1}^{n-1} (Z_{i+1} - Z_i)^2/2(n-1) \tag{7}$$

be used to estimate σ^2. Williams [40] and von Neumann [33, 34] studied the ratio δ^2/s^2, where $s^2=SZ^2/n$ and $Z=Y-\overline{Y}$. Young [42] tabulated significance levels of a linear function of this ratio by use of an Incomplete Beta approximation. Hart [19, 20] tabulated probabilities by use of a series approximation suggested by R. H. Kent. We note that $\delta^2/s^2=2n(1-r_c)/(n-1)$, where

$$r_c = \frac{\frac{1}{2}(Z_1^2 + Z_n^2) + \overset{n-1}{\underset{i=1}{S}} Z_i Z_{i+1}}{\overset{n}{\underset{i=1}{S}} Z_i^2}. \tag{8}$$

T. W. Anderson [14] showed that r_c could be used instead of r_1 to test the hypothesis $\rho = 0$ for Koopmans' model. T. W. Anderson transformed Hart's significance levels [20] to significance levels of r_c.

A non-parametric test for randomness by Wald and Wolfowitz [36] is based on the numerator of r_1' not corrected for the mean. Wallis and Moore [37] developed a series of non-parametric tests based on the signs of differences. Further contributions were made by Rubin [32], Madow [26], Hsu [21], Leipnik [25], Lehmann [23] and Quenouille [29, 30, 31].

If we let ϵ be the error vector and $\sigma^2\alpha$ its covariance matrix, dependent upon $\sigma^2, \rho_1, \rho_2, \cdots, \rho_{n-1}$, Lehmann and Stein [24] have shown that the best test statistic to test the hypothesis that all $\rho_i = 0$ is

$$\frac{\epsilon'\alpha^{-1}\epsilon}{\epsilon' I \epsilon}. \tag{9}$$

Whittle [39] used this method to test the null hypothesis that the data follow a first order moving average against the alternative that they follow an autoregressive scheme of first order, and vice versa.

T. W. Anderson and the author [13] derived the distribution of the circular autocorrelation coefficient for residuals from a fitted Fourier series. Significance levels were found and their use indicated. Exact distributions were possible because of the correspondence between the sine and cosine variables and the λ's in the distribution of the numerator.

Durbin and Watson [17] derived some approximate tests of autocorrelation of the successive residuals in least squares regression with fixed X's. Let the n successive least squares residuals be $Z_1, Z_2, \cdots, Z_n$. Durbin and Watson chose a modification of the von Neumann statistic,

$$d = \frac{\overset{n-1}{\underset{1}{S}} (Z_{i+1} - Z_i)^2}{\overset{n}{\underset{1}{S}} Z_i^2} = \frac{S(\Delta Z)^2}{SZ^2}, \tag{10}$$

to test for the existence of autocorrelation in the errors (ϵ's). We note

$$\frac{\delta^2}{s^2} = \frac{nd}{n-1},$$

so that that $d=2(1-r_c)$, where r_c was T. W. Anderson's statistic [14] to test the hypothesis $\rho=0$ for Koopman's model. It should be emphasized that the original von Neumann and T. W. Anderson statistics do not refer to deviations from a fitted regression; hence the Hart [20] and T. W. Anderson [14] significance levels cannot be used here. However, it would appear reasonable to expect that if we have a large positive autocorrelation, d should be near zero; and for a large negative autocorrelation, d should be near four.

Unfortunately an exact distribution could not be evaluated because the regression variables were not latent roots of the numerator matrix. Hence, only upper and lower bounds of the significance levels (d_U and d_L) could be computed. This was done for 5%, 2.5%, and 1% one-tailed tests, for $n=15$ (1) 40 (5) 100 and for $r=1$ (1) 5. It should be noted that d_U and d_L diverge more as r increases and also as n decreases.

In most cases, the experimenter desires a test of the null hypothesis against the alternative of positive correlation. Hence, one should expect a small value of d when the null hypothesis is false, and we should use the following testing procedure: If the computed value, d, is less than the tabulated value, d^*, the null hypothesis is rejected. On the other hand if the alternative hypothesis is negative correlation, one would expect a value of d near 4 when the null hypothesis is false. In this case we consider $d'=4-d$ and test d' against d^*, as above. Since only upper and lower bounds on the significance levels are available, we proceed as follows:

(i) If d (or d') is less than d_L, reject the null hypothesis.
(ii) If d (or d') is greater than d_U, do not reject.
(iii) If $d_L < d$ (or d') $< d_U$, the test is inconclusive.

If the experimenter wants a two-tailed test, he doubles the significance probability and proceeds as follows:

(i) If d or d' is less than d_L, reject.
(ii) If $d_U < d < 4-d_U$, do not reject.
(iii) Inconclusive otherwise.

An approximate procedure is available for large values of $(n-r-1)$, say greater than 40. In this case, $(1/4)d$ was transformed to a Beta distribution, as Dixon [16] did for r, with parameters p and q, where

$$p + q = \frac{E(d)[4 - E(d)]}{\sigma^2(d)} - 1$$

$$p = \tfrac{1}{2}(p + q)E(d).$$

An approximate test statistic is $F = [p(4-d)]/qd$ with $n_1 = 2q$ and $n_2 = 2p$ degrees of freedom. Or one can use Incomplete Beta tables. Durbin and Watson also present another approximation. Formulas for $E(d)$ and $\sigma^2(d)$ are presented in the 1951 article. Unfortunately, exact significance levels are really needed for small values of $(n-r-1)$, when d_L and d_U tend to be wide apart.

Durbin and Watson [17, 1951] also present methods of testing for autocorrelation with one- and two-way classification data and for curvilinear regression with equally spaced X's.

An example is presented for each of the three types of regression models. Short-cut methods of computing $S(\Delta Z)^2$ are presented for each case. Of course, SZ^2 is simply the error sum of squares. For example, with multiple or curvilinear regression, where $Y - \overline{Y}$ is estimated by

$$\sum b_i(X_i - \overline{X}_i), \qquad \Delta Z = \Delta Y - \sum b_i \Delta X_i.$$

Hence,

$$S(\Delta Z)^2 = S(\Delta Y)^2 + \sum_i \sum_j b_i b_j S(\Delta X_i - \Delta X_j) - 2 \sum_i b_i S(\Delta X_i \Delta Y).$$

Special formulas can be used for curvilinear regression, because of the orthogonal polynomials used in computing the regression coefficients.

Moran [28] presents an exact test of autocorrelation of the residuals, Z_i, when only one predictor is used. He uses the circular autocorrelation coefficient,

$$R_1 = \frac{SZ_i Z_{i+1}}{SZ_i^2},$$

where $Z_{n+1} = Z_1$, and gives formulas for $E(R_1)$ and $\sigma^2(R_1)$.

3. ESTIMATING REGRESSION COEFFICIENTS WHEN THE ERRORS ARE AUTOCORRELATED

To date most of the successful research on autocorrelation has been devoted to the problem of testing for its existence. All too little is known of what to do if the errors actually are autocorrelated. Aitken [1] first showed that if one knew the population covariance matrix for the ϵ's, he could transform the regression model so that the method

of least squares would give efficient estimates of the β's. If the covariance matrix of the ϵ's is $\alpha\sigma^2$ and the regression model (in matrix form) is $Y = X\beta + \epsilon$, we premultiply this regression model by the non-singular matrix H, where

$$H\alpha H' = I,$$

and I is the $n \times n$ identity matrix. But even if α were known, the solution for H might be very difficult. However, if the ϵ's follow a first order autoregressive process with autocorrelation ρ and variance $\sigma^2/(1-\rho^2)$, the transformation is quite simple:

$$\epsilon_1^* = \sqrt{1-\rho^2}\,\epsilon_1, \quad \epsilon_i^* = \epsilon_i - \rho\epsilon_{i-1}, \quad \text{for} \quad 1 < i \leqq n.$$

The transformations for higher order autoregressive processes and for moving average processes are more complicated. A good explanation of this is given by Watson [54].

Allowing for the difficulty of making the transformation if α is known, the major defect is the lack of knowledge regarding the true value of α. Most time series are too short to enable one to derive good estimates of the parameters in α, or even to determine the type of process which is operating. A recent attempt to bypass the transformation problem when the ϵ's follow an autoregressive process was made by Champernowne [44]. He assumes that the model for the ϵ's is

$$\sum_{s=0}^{m} \gamma_s(\epsilon_{t-s} - \alpha) = \delta_t,$$

where the δ's are assumed normally and independently distributed with zero mean and variance σ^2. Champernowne presents the following results:

(i) Assuming the γ's are known, $\hat{\alpha}$ was determined as a weighted mean and $\hat{\sigma}^2$ as a weighted quadratic function of observed values of the ϵ's.
(ii) Assuming the γ's are known, estimates of and confidence limits for the regression coefficients are derived, both with α known and α unknown.
(iii) The results in (ii) are derived for $\alpha = 0$.
(iv) If the γ's are not known, the least-sqares estimates of the regression coefficients are not linear functions of the observed Y's and X's; hence, the usual χ^2 distribution theory does not hold exactly. A method involving the application of Bayes' Theorem was used in this case.
(v) A brief discussion is given of these problems when the X's also have disturbances.

Cochrane and Orcutt [46] indicate three principal reasons that the ϵ's in economic time series models tend to be positively autocorrelated:

(i) Faulty choice of the form of the regression model.
(ii) Omission of important variables from the model.
(iii) Use of incorrect variables or poor data.

They analyzed the sample residuals for a number of econometric studies, and found many significant autocorrelations, using von Neumann's statistic, δ^2/s^2. As indicated earlier, this statistic does not take account of the added correlation of the estimated residuals resulting from the necessity of estimating the regression coefficients; this defect becomes worse as the number of X's increase. In addition δ^2/s^2 does not take account of the autoregressive nature of many X-variables.

Cochrane and Orcutt also conducted some empirical sampling experiments to indicate the effects of autoregressive error processes on least squares regression analysis, with the following indicated results:

(i) The sample residuals tend to be biased towards randomness.
(ii) The variance of least squares estimates of the regression coefficients are very large if the errors are highly autocorrelated (in their example, $\rho \geq .8$).
(iii) If the autocorrelations could be reduced to $\rho < .3$ or perhaps even $\rho < .5$, by use of a simple transformation, these variances appear to be close to those with random errors.
(iv) The removal of trend seems to be a crude but effective transformation in many cases.
(v) If sample residuals are used to estimate the error variance, σ^2, this estimate will be too small if the errors are positively correlated. This result can be proven exactly, see for example, Cochran [45].

Cochrane and Orcutt state that for many economic variables, it is a simple and practical procedure to analyze the first differences of the various series. If the original regression equation is

$$Y_t = \beta_0 + \sum_{i=1}^{r} \beta_i X_{it} + \epsilon_t,$$

the transformed equation will be

$$\Delta Y_t = \sum \beta_i \Delta X_{it} + \Delta \epsilon_t,$$

where $\Delta Z_t = Z_{t-1} - Z_t$. This would be the exact transformation for $\rho = 1$ in a first order autoregressive model, except that ρ must be less than 1 in order to avoid an explosive situation. However, the transformation should be reasonably good if ρ is near 1, and it is certainly very simple. If the sample residuals, after transforming the variables in this manner, are still highly autocorrelated, one might use the estimated autocorrelation coefficients to try a new transformation.

Stone [52] used the method of first differences advocated by Cochrane and Orcutt [46] to reanalyze his market demand data [51]. Stone uses the von Neumann statistic with the sample residuals to test for autocorrelation in the errors. He found the average autocorrelation for 13 analyses highly significant before transforming and almost equal to its expectation after transforming. It was interesting to note that the two sets of regression coefficients were not materially different.

Watson [54] has investigated the efficiencies and estimated variances of least-squares estimates of regression coefficients for fixed X's and tests of hypotheses concerning them, when an incorrect transforming model is used. General solutions of the following type are presented: bounds on the bias of the estimated variance, lower bound to the efficiency of the estimates of regression coefficients and some bounds on the significance points of the t- and F-tests. He then discusses the following special types of incorrect transformations:

(i) *Assumed and true error processes are both autoregressive.*
  (a) *Both are first order but an incorrect ρ_1 is used.* The greatest bias to the estimated variance is a downward bias when ρ is underestimated. This offers some justification for the use of the first difference transformation, which overestimates ρ. ρ is generally underestimated from sample residuals. However, we note low efficiencies of estimates of regression coefficients when ρ is overestimated unless ρ is nearly 1.
  (b) *True process is second order and assumed process is first order.* Results depend on how accurately one knows ρ_1 and on the magnitude of ρ_2.

(ii) *Assumed and true error processes are both moving averages.*
  (a) *Both are first order with incorrect ρ_1 used.* Results in (i) are reversed.
  (b) *True process is second order and assumed process is first order.* Indications are that an incorrect order is more serious for a moving average than for an autoregressive process.

(iii) *Assumed process is first order autoregressive and true process is first order moving average.* Even when ρ_1 is estimated correctly the bias in the variance can be appreciable and the efficiency quite low.

In all cases the true probabilities for 5% significance levels may be considerably different, the bounds being of the order of less than 1% to over 10% in many cases of what would appear to be only mildly inaccurate estimates.

Watson is rather pessimistic regarding the use of transforming devices to remove the effect of autocorrelation in least square analysis of time series data. However, he believes that more investigations need to be made of correlograms of residuals to see if a good analysis can be constructed on the basis of these correlograms. Quenouille [48] presents a test of the hypothesis that a sample was drawn from an auto-

regressive scheme of specified order and Wold [55] did the same for a moving average process. Similar tests are given by Bartlett and Diananda [43] and Walker [53]. However more efficient methods are needed, and especially we need to determine the proper process and order. After all, as Watson [54] remarks, one must use some kind of an analysis, and it is the duty of the statistician to find a good method, even if it is not the correct one.

A series of articles sponsored by the Indian Statistical Institute [47] describe the results of using empirical sampling methods to evaluate the usefulness of the Wold [55] and Quenouille [48] large sample tests for short series. Matthai and Kannan considered three different moving average models and S. R. Rao and Som two autoregressive models. Series of length 15 and 35 were used. It was shown that both large sample tests gave far too many significant results for the short series used. Quenouille's test showed that a second order autoregressive model would not fit third order moving average data; however, Wold's test indicated that a third order moving average model could be used even if the data were second order autoregressive. This may indicate that a moving average model of high order is more likely to represent a given set of data than is an autoregressive model. Or it may indicate that Quenouille's test is more powerful than Wold's in indicating the correct process. It was interesting to note that in both studies the correlogram was well estimated, if one knew the correct process. The third paper, by C. R. Rao, presents a sequential procedure for determining the number of sample autocorrelation coefficients needed to estimate the correlogram. Rao advocates the use of likelihood to discriminate between several possible models to represent a given set of data.

Sastry [49] used the above models and data to investigate the small sample bias in the estimates of the autocorrelation coefficients. He first compared definitions (2) and (3) and concluded that (2) was superior. However (3) is better for small lags and is certainly much easier to compute. In general small sample estimates have large biases, even for series of 100. The size of the bias depends on the type of model (it was much less for a second order autoregressive model than for the other four models) and on the values of the parameters in the model.

Sastry [49] also considers some theoretical results for comparing two series of autocorrelated variates, x and y. He presents the expected values of the means, variances, variances of the means, and covariances of x and y, and some higher moments for normal variates. He proposes this new statistic to test the hypothesis that $E(x)=\mu$:

$$t' = \frac{\sqrt{\left\{(n-1) - \frac{2}{n}\sum_k (n-k)\rho_k\right\} n(\bar{x}-\mu)^2}}{\sqrt{\left\{1 + \frac{2}{n}\sum_k (n-k)\rho_k\right\} S(x-\bar{x})^2}}$$

with

$$f' = \left\{n - 1 - \frac{2}{n}\sum_k (n-k)\rho_k\right\}$$

degrees of freedom. Sastry does not indicate how useful t' will be when ρ_k must be estimated from the data. One can surely see that relatively unbiased estimates need to be obtained. And, most important for regression analysis, he presents the expected values and variances of estimates of the parameters in $E(y)=\alpha+\beta x$.

4. FURTHER COMMENTS ON AUTOREGRESSIVE MODELS

Although the main topic of this paper is a discussion of regression analysis with fixed X's, some references on the use of autoregressive models will be included. These models were first discussed by Yule and have been used extensively by economists. The regression coefficients in these models are functions of the autocorrelation coefficients. Hurwicz [58] shows that least squares estimates of the parameters are biased in small samples. As indicated previously, Mann and Wald [6] showed that this bias approached zero as the sample size increased. Only large sample least-squares variances and covariances of the estimates of the parameters are available; hence, confidence limits for the parameters and predicted values are available only for large samples. Tintner [7] presents an example for a third-order process. Kendall [5, 59] gives further information on the use of least squares to estimate the parameters.

Bartlett [2] presents a method of estimation based on the concept of a continuous rather than a discrete process. Ghurye [57] has developed a method of using more of the autocorrelation coefficients in estimating the parameters. He introduces a superposed variation for each operation, so that the model is (assuming $\beta_0=0$):

$$(Y_t + \eta_t) = \sum_i \beta_i(Y_{t-i} + \eta_{t-i}) + \epsilon_t.$$

An extensive study of autoregressive analysis is presented by Orcutt [60].

Das [56] uses empirical sampling methods to measure the goodness of least squares methods for three equation economic models, in which one equation is quantity in terms of present prices and the others involve only lagged variables as predictors.

5. OMITTED TOPICS

The following topics, of importance in the analysis of time series, have not been discussed in detail.

(i) *The estimation of parameters in a multi-equation system.* For a discussion of this procedure, see for example Koopmans [62] and Klein [61]. An article by Orcutt and Cochrane [64] presents an empirical sampling study of the adverse effect of autocorrelation on the estimates of structural parameters in a multi-equation model. They concluded that, "Unless it it possible to specify something about the intercorrelations of the error terms in a set of relations and to choose approximately the correct autoregressive transformation, a certain amount of skepticism is justified concerning the possibility of estimating structural parameters from aggregative time series of only twenty observations."

(ii) *Comparing two time series.* See Bartlett [15, 2], Orcutt and James [65] and Moran [63].

6. SUMMARY

Much research has been devoted to the distributions of various statistics used to test for the existence of autocorrelation of successive observations. Others have studied the problem of estimating parameters in various stochastic processes, such as autoregressive and moving average processes. A summary of this research is given in this paper.

Only recently has research been extended to the problem of testing for the existence of autocorrelated errors in regression models, such as

$$Y_t = \beta_0 + \sum_{i=1}^{r} \beta_i X_{it} + \epsilon_t, \qquad t = 1, 2, \cdots, n,$$

where the X's are fixed predictors and the ϵ's are normally distributed with equal variance. Durbin and Watson [17] present upper and lower bounds on the significance levels for making such tests. Moran [28] presents an exact test for $r=1$.

Too little information is available on the proper methods of estimating the β's when the ϵ's are autocorrelated. Aitken [1] indicated the exact method of transforming the regression variables when the autocorrelations were known. Champernowne [44] added to this general theory and presented a Bayesian method when the autocorrelations were not known.

Cochrane and Orcutt [46] used empirical sampling methods to indicate the effects of autocorrelated errors on the estimates of error and the β's. They showed that, in many cases, first differences of the Y's and X's would have a relatively uncorrelated error process. A series of articles in Sankhyā [47] have also used empirical sampling to indicate the large biases in testing and estimation procedures with small samples.

Watson [54] has shown the seriousness of using the wrong type of error process and incorrect estimates of the autocorrelations in transforming the regression variables. He concludes that the most fruitful research seems to be in utilizing more efficiently the estimates of the autocorrelations.

BIBLIOGRAPHY

1. *General*

[1] Aitken, A. C., "On least squares and linear combinations of observations," *Proceedings of the Royal Society of Edinburgh*, 55 (1935) 42–48.

[2] Bartlett, M. S., "On the theoretical specification and sampling properties of autocorrelated time series." *Journal of the Royal Statistical Society Supplement*, 8 (1946) 27–41.

[3] Cunningham, L. B. C. and Hynd, W. R. B., "Random processes in problems in air welfare," *Journal of the Royal Statistical Society Supplement*, 8 (1946) 62–85.

[4] Foster, G. A. R., "Some instruments for the analysis of time series and their applications to textile research," *Journal of the Royal Statistical Society Supplement*, 7 (1946) 42–61.

[5] Kendall, M. G., *The advanced theory of statistics*, Charles Griffin and Company, Limited, London, 1948.

[6] Mann, H. B. and Wald, A., "On the statistical treatment of linear stochastic difference equations," *Econometrica*, 11 (1943) 173–220.

[7] Tintner, G., *Econometrics*, John Wiley and Sons, Incorporated, New York, 1952.

[8] Wold, H., *A study in the analysis of stationary time series*, Almquist and Wiksells, Uppsala, 1938.

[9] Yule, G. U., "On the time-correlation problem," *Journal of the Royal Statistical Society*, 84 (1921) 496–537.

[10] Yule, G. U., "Why do we sometimes get nonsense correlations between time series?", *Journal of the Royal Statistical Society*, 89 (1926) 1–64.

2. *Tests of Significance for Autocorrelation*

[11] Anderson, R. L., "Serial correlation in the analysis of time series," Unpublished Thesis, Library, Iowa State College, 1941.

[12] Anderson, R. L., "Distribution of the serial correlation coefficient," *Annals of Mathematical Statistics*, 13 (1942) 1–13.

[13] Anderson, R. L. and Anderson, T. W., "Distribution of the circular serial correlation coefficient for residuals from fitted Fourier series," *Annals of Mathematical Statistics*, 21 (1950) 59–81.

[14] Anderson, T. W., "On the theory of testing serial correlation," *Skandinavisk Aktuarietidskrift*, 31 (1948) 88–116.

[15] Bartlett, M. S., "Some aspects of the time-correlation problem in regard to tests of significance," *Journal of the Royal Statistical Society*, 98 (1935) 536–43.

[16] Dixon, W. J., "Further contributions to the problem of serial correlation," *Annals of Mathematical Statistics*, 15 (1944) 119–44.

[17] Durbin, J. and Watson, G. S., "Testing for serial correlation in least squares regression," *Biometrika*, 37 (1950) 409–28; 38 (1951) 159–78.

[18] Fisher, R. A., "Tests of significance in harmonic analysis," *Proceedings of the Royal Society Series A*, 125 (1929) 54–59.

[19] Hart, B. I., "Tabulation of the probabilities of the ratio of the mean square successive difference to the variance," *Annals of Mathematical Statistics*, 13 (1942a) 207–14.

[20] Hart, B. I., "Significance levels for the ratio of the mean square successive difference to the variance," *Annals of Mathematical Statistics*, 13 (1942b) 445–47.

[21] Hsu, P. L., "On the asymptotic distribution of certain statistics used in testing the independence between successive observations from a normal population," *Annals of Mathematical Statistics*, 17 (1946) 350–54.

[22] Koopmans, T., "Serial correlation and quadratic forms in normal variables," *Annals of Mathematical Statistics*, 13 (1942) 14–23.

[23] Lehmann, E. L., "On optimum tests of composite hypotheses with one constraint," *Annals of Mathematical Statistics*, 18 (1947) 473–94.

[24] Lehmann, E. L. and Stein, C., "Most powerful tests of composite hypotheses I Normal distributions," *Annals of Mathematical Statistics*, 19 (1948) 495–516.

[25] Leipnik, R. B., "Distribution of the serial correlation coefficient in a circularly correlated universe," *Annals of Mathematical Statistics*, 18 (1947) 80–87.

[26] Madow, W. G., "Note on the distribution of the serial correlation coefficient," *Annals of Mathematical Statistics*, 16 (1945) 308–10.

[27] Moore, G. H. and Wallis, W. A., "Time series significance tests based on signs of differences," *Journal of the American Statistical Association*, 38 (1943) 153–64.

[28] Moran, P. A. P., "A test for the serial independence of residuals," *Biometrika*, 37 (1950) 178–81.

[29] Quenouille, M. H., "Some results in the testing of the serial correlation coefficient," *Biometrika*, 35 (1948) 261–67.

[30] Quenouille, M. H., "Approximate tests of correlation in time series," *Journal of the Royal Statistical Society Supplement*, 11 (1949a) 68–84.

[31] Quenouille, M. H., "The joint distribution of serial correlation coefficients," *Annals of Mathematical Statistics*, 20 (1949b) 561–71.

[32] Rubin, H., "On the distribution of the serial correlation coefficient," *Annals of Mathematical Statistics*, 16 (1945) 211–15.

[33] von Neumann, J., "Distribution of the ratio of the mean square successive difference to the variance," *Annals of Mathematical Statistics*, 12 (1941b) 367–95.

[34] von Neumann, J., "A further remark on the distribution of the ratio of the

mean square successive difference to the variance," *Annals of Mathematical Statistics*, 13 (1942) 86–88.

[35] von Neumann, J., Kent, R. H., Bellinson, H. R., and Hart, B. I., "The mean square successive difference," *Annals of Mathematical Statistics*, 12 (1941a) 153–62.

[36] Wald, A. and Wolfowitz, J., "An exact test for randomness in the non-parametric case based on serial correlation," *Annals of Mathematical Statistics*, 14 (1943) 378–88.

[37] Wallis, W. A. and Moore, G. H., "A significance test for time series analysis," *Journal of the American Statistical Association*, 36 (1941) 401–09.

[38] Watson, G. S. and Durbin, J., "Exact tests of serial correlation using non-circular statistics," *Annals of Mathematical Statistics*, 22 (1951) 446–51.

[39] Whittle, P., *Hypothesis testing in time series*, Almqvist and Wiksells Boktryckeri A B, Uppsala, (1951).

[40] Williams, J. D., "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe," *Annals of Mathematical Statistics*, 12 (1941) 239–51.

[41] Wilson, E. B., "Periodogram of American business activity," *Quarterly Journal of Economics*, 48 (1934) 375–417.

[42] Young, L. C., "On randomness in ordered sequences," *Annals of Mathematical Statistics*, 12 (1941) 293–300.

3. *Estimating Regression Coefficients When the Errors are Autocorrelated*

[43] Bartlett, M. S. and Diananda, P. H., "Extensions of Quenouille's test for autoregressive schemes," *Journal of the Royal Statistical Society Series B*, 12 (1950) 108–15.

[44] Champernowne, D. G., "Sampling theory applied to autoregressive sequences," *Journal of the Royal Statistical Society Supplement*, 10 (1948) 204–42.

[45] Cochran, W. G., "Some consequences when the assumptions for the analysis of variance are not satisfied," *Biometrics*, 3 (1947) 32–35.

[46] Cochrane, D. and Orcutt, G. H., "A sampling study of the merits of autoregressive and reduced form transformations in regression analysis," *Journal of the American Statistical Association*, 44 (1949) 356–72.

[47] Indian Statistical Institute, "The applicability of large sample tests for moving average and autoregressive schemes to series of short length—an experimental study." *Sankhyā*, 11 (1951) 217–72.
   1. Matthai, A. and M. B. Kannan, "Moving averages."
   2. Rao, S. R. and R. K. Som, "Autoregressive series."
   3. Rao, C. R., "The discriminant function approach in the classification of time series."

[48] Quenouille, M. H., "A large-sample test for the goodness of fit of autoregressive schemes," *Journal of the Royal Statistical Society*, 110 (1948) 123–29.

[49] Sastry, A. S. R., "Bias in estimation of serial correlation coefficients," *Sankhyā*, 11 (1951a) 281–96.

[50] Sastry, A. S. R., "Some moments of moment statistics and their use in tests of significance in autocorrelated series," *Sankhyā*, 11 (1951b) 297–308.

[51] Stone, R., "The analysis of market demand," *Journal of the Royal Statistical Society*, 108 (1945) 1–98.

[52] Stone, R., "The analysis of market demand," *Review of the International Statistical Institute*, 10 (1948) 1–13.

[53] Walker, A. M., "Notes on a generalization of the large sample goodness of fit test for linear autoregression schemes," *Journal of the Royal Statistical Series B*, 12 (1950) 102–07.

[54] Watson, G. S., *Serial Correlation in Regression Analysis*, Unpublished Thesis, Library, North Carolina State College, also Institute of Statistics, Mimeo Series No. 49, (1951).

[55] Wold, H., "A large-sample test for moving averages," *Journal of the Royal Statistical Society, Series B*, 11 (1949) 297–305.

4. *Autoregressive Models*

[56] Das, A. C., "On the estimation of parameters in a recursive system," *Sankhyā*, 11 (1951) 273–80.

[57] Ghurye, S. G., "A method of estimating the parameters of an autoregressive time series," *Biometrika*, 37 (1950) 173–78.

[58] Hurwicz, L., "Least squares bias in time series," Chap. 15 in Koopmans, T. C., *Statistical Inference in Dynamic Economic Models*, John Wiley and Sons, Incorporated, New York, 1950.

[59] Kendall, M. G., "On autoregressive time series," *Biometrika*, 33 (1944) 105–22.

[60] Orcutt, G. H., "A study of the autoregressive anture of the time series used for Tinbergen's model of the economic system of the United States, 1919–32," *Journal of the Royal Statistical Society Supplement*, 10 (1948) 1–53.

5. *Other Topics*

[61] Klein, L. R., "The use of econometric models as a guide to economic policy," *Econometrica*, 15 (1947) 111–51.

[62] Koopmans, T. C., *Statistical Inference in Dynamic Economic Models*, John Wiley and Sons, Incorporated, New York, (1950).

[63] Moran, P. A. P., "Some theorems on time series," *Biometrika*, 34 (1947) 281–91; 35 (1948) 255–60.

[64] Orcutt, G. H. and Cochrane, D., "A sampling study of the merits of autoregressive and reduced form transformations in regression analysis," *Journal of the American Statistical Association*, 44 (1949) 356–72.

[65] Orcutt, G. H. and James, S. F., "Testing the significance of correlation between time series," *Biometrika*, 35 (1949) 397–413.

JOINT CONFIDENCE REGIONS FOR MULTIPLE REGRESSION COEFFICIENTS

DAVID DURAND
National Bureau of Economic Research

More and more statisticians are coming to realize that conventional confidence intervals are not strictly applicable to problems requiring the estimation of several parameters. In multiple regression a conventional interval may be correctly determined for one, and usually only one, of the regression coefficients. Ordinarily, however, the statistician wants a measure of accuracy for each of his coefficients, but if he obtains these in the form of conventional confidence intervals, he usually commits a fallacy. Here we discuss the nature of this fallacy and a possible remedy through the use of a joint confidence region.

1. MULTIPLE CONFIDENCE STATEMENTS

In classical multiple regression it is assumed that the dependent variate Y is normally distributed with constant variance about a linear function

$$b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k \tag{1.1}$$

in which the coefficients b_i as well as the variance σ^2 are unknown parameters. From a set of $n>k+1$ error free and linearly independent observations on the X_i's and n corresponding values for Y, one obtains the estimates $\hat{b}_i$ and $\hat{\sigma}^2$—here understood to be maximum likelihood estimates. Then, the theory of confidence provides criteria for judging the accuracy of these estimates.

After deriving the k-variable regression function (1.1) many statisticians would want exactly $k+2$ confidence statements—one for each of the $k+1$ b_is and one for σ^2. But this rule is far from general. Personal preference or the requirements of the problem may dictate 1, 2, $\cdots$ $k+2$, or even more statements. Since the theory of confidence permits statements for linear combinations of the type

$$c_0b_0 + c_1b_1 + c_2b_2 + \cdots + c_kb_k,$$

there is literally no limit to the number of confidence statements that can be considered.

The conventional form of confidence interval for a partial regression coefficient is determined by the relation

$$(1.2)\qquad \frac{|\,\hat{b}_p - b_p|}{\sqrt{a^{pp}n\hat{\sigma}^2/(n-k-1)}} \leq t_\alpha$$

where t_α is the upper $\frac{1}{2}\alpha$ point of 'Student's' ratio for $n-k-1$ degrees of freedom, and a^{pp} is the element corresponding to a_{pp} in the inverse of the matrix $\|a_{ij}\| = \|\sum_{t=1}^{n}(X_{it}-\overline{X}_{it})(X_{jt}-\overline{X}_{jt})\|$. Given a value for α, say .05, (1.2) determines an interval that will cover the true parameter value b_p with probability $1-\alpha$.[1] But although these conventional intervals are entirely valid when properly applied and interpreted, there are two basic fallacies or improprieties that frequently arise in practical work. The first consists in deciding after the experiment what confidence statements to make. The second consists in making several individual statements at level $1-\alpha$ in a way that implies a joint statement at the same level.

Concerning the first fallacy, it is common practice in statistical studies in general to go over the data with a fine-toothed comb, to apply a battery of significance tests, and then to select a relatively few conclusions that seem particularly noteworthy. In regression studies it is common to experiment with several equations before selecting one for presentation. This might consist in calculating a regression equation with five variables, discovering that two of the coefficients do not differ significantly from zero, and then recomputing the equation with three variables. But procedures such as these may introduce bias, as may be seen from an extreme example. Suppose that the true regression coefficients were all zero throughout a series of experiments, and suppose that the experimenter made a practice of presenting only regression equations with coefficients significantly different from zero at the level α. Then, if he made confidence statements for these coefficients, his probability of being right would not be $1-\alpha$ at all, but zero; for he would make a statement only when the parameter value zero lay outside the confidence interval.

The second common fallacy in using conventional confidence intervals is the implied joint statement. In a two-variable problem—such as that presented in Section 3—we might make the following three conventional statements at the 95 per cent level:

---

[1] Perhaps it is necessary, for the record, to say a word about the meaning of "probability" in this context. Once the experiment has been performed, the interval either does or does not cover the point b_p, and the probability is therefore 1 or 0. Before the experiment, however, we may argue that the outcome is uncertain and that the probability is properly $1-\alpha$. And before performing a series of experiments, we may argue that $1-\alpha$ is the proportion of projected confidence statements that will be correct in the long run.

$$
\begin{array}{rcl}
-.04 \leqq & b_1 & \leqq .44 \\
.52 \leqq & b_2 & \leqq .98 \\
.91 \leqq & b_1 + b_2 & \leqq .99.
\end{array} \tag{1.3}
$$

The first statement locates the parameter point (b_1, b_2) within an infinite band bounded by two lines parallel to the b_2 axis in two-dimensional parameter space (Figure 2); the second locates the same point in a band between lines parallel to the b_1 axis; and the third locates the point between two parallel sloping lines. But the three statements, taken all together, locate the point within the hexagon formed by the intersection of the three bands. So, if the confidence level of the individual statements is $1-\alpha=.95$, the level of the joint statement is demonstrably less.

The actual confidence level of a set of statements like (1.3) can be readily obtained for one or two special cases, the most obvious of which occurs when, in a k-variable problem, the cross-products $a_{ij}(i\neq j) = \sum_{t=1}^{n}(X_{it}-\overline{X}_i)(X_{jt}-\overline{X}_j)$ are all zero and the variance σ^2 is known. Then, the sampling distributions of the individual $\hat{b}_i$s are all independent, and the joint confidence level for any subset containing exactly m of these coefficients is therefore $(1-\alpha)^m$. Another special case, involving differences between means in the analysis of variance, has been discussed by Tukey [11]. But in general the joint confidence level of statements like (1.3) is not readily obtained, though a lower bound can be obtained as shown in Section 5.

As an alternative to calculating the confidence level of joint statements like (1.3), Scheffé [9], Roy and Bose [8], and the author propose to define an infinite set of intervals whose totality is equivalent to a joint confidence ellipse at the level $1-\alpha$. Although confidence ellipses have been understood in theory for some time, they have received little practical application—notwithstanding an important econometric example by Haavelmo [4]. Possibly, the unpopularity of the ellipse is due partly to the difficulty of representing it graphically—except in two dimensional examples, like Haavelmo's—but this is largely obviated by the proposed technique of substituting an infinite set of intervals.

The use of the ellipse with an infinite set of intervals has two distinct advantages. First, the calculations required are no more difficult than those required in the conventional approach. Second, if a finite or infinite subset of the intervals is chosen in any way whatsoever, before or after the experiment, the confidence level cannot be less than $1-\alpha$; thus the fallacy of choosing statements after the experiment is avoided. At the same time, this approach has one drawback; whenever interest is

limited to a finite subset of intervals that can be specified in advance, the confidence level will actually exceed $1-\alpha$, and the intervals will therefore be larger than necessary. For example, Tukey's method for contrasting means in the analysis of variance allows for just $\frac{1}{2}k(k-1)$ specified contrasts among k means; and for $k>2$, as shown by Scheffé [9], Tukey's intervals are smaller than those derived from the joint ellipse, and the difference increases as k increases. However, it should be remembered that the joint ellipse is proposed primarily for investigations where the specification of questions in advance is not convenient. The next section presents an example characterized by a proliferation of possible questions and great need for flexibility, and it is for problems of this sort that the joint ellipse, with its infinite set of intervals, is ideally suited.

2. AN EXAMPLE

In an exploratory cross-section study of seventeen New York bank stocks for February 1951, the dependent variate, log P, had an estimated variance $\hat{\sigma}^2=.0006863$ about the estimated regression plane

$$\log P = .037 + .65 \log C - .95 \log S + .26 \log D, \tag{2.1}$$

TABLE 1

MEANS AND SUMS OF SQUARES AND PRODUCTS ABOUT THE MEANS FOR REGRESSION ANALYSIS OF 17 NEW YORK CITY BANK STOCKS, FEBRUARY 1951

Note: The variables in this example were all expressed as common logarithms with 2-place mantissas. The sums of squares and products were rounded to 5 decimals for use in calculations.

	X_1 Log Capital (year end 1950, unit: \$10,000)	X_2 Log Shares (February 1951, unit: 10,000)	X_3 Log Dividends (total disbursements 1950, unit: \$100)	Y Log Price (February 1951, unit: \$1.00)
Means	3.9288	1.9353	4.5206	1.9512
Sums of Squares and Products about the Means				
X_1	3.25538	3.51281	3.42601	− .30838
X_2		7.25942	3.55875	−3.65961
X_3			3.69849	− .16441
Y				3.23758

which then lead to the exponential form

$$P = 1.09C^{.65}S^{-.95}D^{.28}. \tag{2.2}$$

In these equations, P is market price (end of February 1951), C is total capital funds (end of 1950 in units of \$10,000), S is total number of shares outstanding (end of February 1951 in units of 10,000), and D is total dividend disbursements (during 1950 in units of \$100). The means and the a_{ij} matrix for the logarithmic variables are given in Table 1. The forward Doolittle solution, which will be needed in the subsequent discussion, is given in Table 2.

TABLE 2

FORWARD DOOLITTLE SOLUTION FOR REGRESSION COEFFICIENTS IN EQUATION (2.1)

Note: Although entries have been rounded to 4 decimals for illustration here, the original matrix (see Note, Table 1) was obtained to 5 decimals, and subsequent calculations were carried to 8 or 9 decimals.

X_1 Log Capital (\$10,000)	X_2 Log Shares (10,000)	X_3 Log Dividends (\$100)	Y Log Price	Check Sum
3.2554	*3.1528*	*3.4260*	.3084	10.5026
−1.0000	−1.0791	−1.0524	− .0947	−3.2262
3.5128	7.2594	3.5538	3.6596	17.9856
−3.5128	−3.7906	−3.6969	− .3328	−11.3331
	3.4688	− *.1432*	3.3268	6.6525
	−1.0000	.0413	− .9591	−1.9178
3.4260	3.5538	3.6985	.1644	10.8427
−3.4260	−3.6969	−3.6056	− .3245	−11.0531
	.1432	− .0059	.1373	.2746
		.0870	− .0228	.0642
		−1.0000	.2622	− .7378

Though the form of equations (2.1) and (2.2) is convenient for computation and meets the needs of an estimating equation, it does not adequately describe the structure of the bank stock market. For this, one wants to relate stock prices to such variables as book value and dividends per share. However, a simple linear transformation on the logarithms of the independent variables in (2.1) and (2.2)—namely,

$$\log C = \log C$$
$$\log C/S = \log C - \log S$$
$$\log D/S = \log D - \log S$$

produces a new equation

$$\text{(2.3)} \qquad P = 1.09C^{-.04}(C/S)^{.69}(D/S)^{.26},$$

where the transformed independent variables are total capital (which indicates size of bank), capital per share (book value), and dividends per share. In this transformed equation the new coefficients (or exponents) are all linear combinations of the original coefficients, thus:

$$-.04 = .65 + .26 - .95$$
$$.69 = .95 - .26.$$

The variance, $\hat{\sigma}^2 = .0006863$, is unaffected.

At the time this regression analysis was performed, most bank stocks were selling at substantial discounts from book value—a fact that worried many financiers. Again, suitable equations for studying discounts can be derived from (2.2) by other linear transformations on the logarithmic variables. One possible equation is

$$\text{(2.4)} \qquad P/B = 1.09B^{-.09}(D/C)^{.26}S^{-.04}$$

where $B = C/S$ represents book value. Here, as in (2.3), the new coefficients are all linear combinations of the coefficients of (2.1), and the variance remains unchanged.

Thus, by the nature of this problem, it is possible to start with a basic equation, (2.1) or (2.2), and to derive from this by suitable transformations a series of special purpose equations. This process, however, multiplies the number of regression coefficients and combinations for which confidence statements are required. In addition to the constant term, which is not affected by the illustrated transformations, equations (2.2), (2.3), and (2.4) contain five different regression coefficients, and if all the ramifications of the problem were to be explored, more equations and coefficients would undoubtedly arise. Moreover, it should be realized that this example was artificially cut down for simplicity in presentation. A systematic study of bank stock prices should contain anywhere from five to ten basic variables and upwards of twenty-five transformed variables.

To avoid the fallacies of multiple statements in this problem, where so many statements are possible, a joint confidence ellipse will be

determined for b_1, b_2, and b_3. Although the constant term b_0 could be included in the ellipse (see Section 6), this particular problem is primarily concerned with the structure of the market as indicated by b_1, b_2, and b_3—not with the general level of the market as reflected by b_0. However, a statement may be desired for the variance σ^2, and this can be obtained in the conventional manner provided no joint relationship is thus implied for σ^2 and the b_i's.

3. THE JOINT CONFIDENCE ELLIPSOID

A joint confidence region for the k regression coefficients obtained when a single dependent variate is regressed upon k independent variables $X_1, X_2, \cdots, X_k$ is given by the ellipsoid

$$F_\alpha(k, n-k-1) = \frac{(n-k-1)\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}(\hat{b}_i - b_i)(\hat{b}_j - b_j)}{kn\hat{\sigma}^2}, \tag{3.1}$$

where $F_\alpha(k, n-k-1)$ is the upper α point of the F-distribution for k and $n-k-1$ degrees of freedom, n is the number of observations, $a_{ij} = \sum_{\xi=1}^{n}(X_{i\xi}-\overline{X})(X_{j\xi}-\overline{X})$, $\hat{b}_i$ is the maximum likelihood estimate of the true regression coefficient b_i, and $\hat{\sigma}^2$ is the maximum likelihood estimate of the variance.[3] To apply (3.1) a value of α is chosen, say .05, and the single statement is made with probability $1-\alpha=.95$ that the parameter point $b_1, b_2, \cdots, b_k$ lies within the ellipsoid.

To illustrate a joint confidence region graphically, the prices of the seventeen New York City bank stocks were regressed on the two variables dividends per share, D/S, and book value, C/S; and the following resulted, with variables now expressed in dollars:

$$P = 2.15(D/S)^{.20}(C/S)^{.75}. \tag{3.2}$$

A 95 per cent ellipse was then determined by inserting in (3.1) the appropriate numerical values, including $F_{.05}(2, 14)=3.739$ and $\hat{\sigma}^2 = .0008806$. A graph is shown in Figure 1. Strictly speaking, this confidence region applies only to points in two-dimensional parameter space—that is, to paired values of b_1 and b_2. Thus the combination $b_1=.30$ and $b_2=.65$ lies within the ellipse and is admissible, whereas the combination $b_1=.15$ and $b_2=.70$ lies outside and is inadmissible. It will be noticed immediately, however, that for certain values of b_1—

[3] For derivation of (3.1) see Wilks [12, Sec. 8.3]. In following the derivation, one must remember that Wilks makes one of the X_i's, say X_1, arbitrarily equal to unity; hence b_1 in his notation is the same as b_0 herein, and k in his notation is the same as $k+1$ herein.

namely, those less than $-.09$ or greater than $.49$—all points lie outside the ellipse. Likewise, for b_2 less than $.45$ or greater than 1.05, all points lie outside. Thus, we obtain the intervals

$$(3.3)\qquad \begin{aligned} -.09 \leq b_1 &\leq .49 \\ .45 \leq b_2 &\leq 1.05, \end{aligned}$$

each of which corresponds to a confidence level exceeding $1-\alpha$. Although intervals of this type have been called confidence intervals

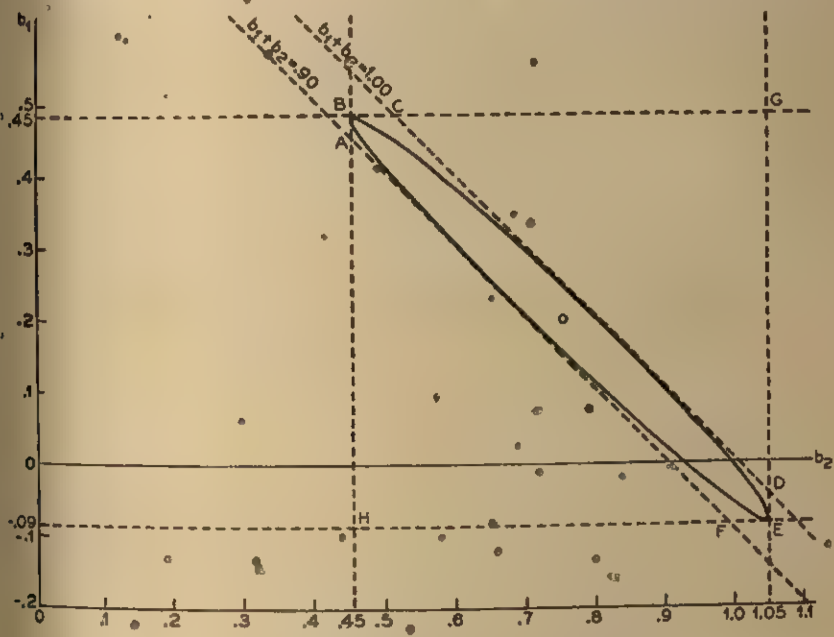


FIG. 1. 95 per cent joint confidence ellipse and subsidiary intervals for (3.2).

they are here referred to as "subsidiary intervals" (subsidiary to the ellipse) in order to distinguish them from the conventional intervals.

Similar subsidiary intervals can be derived, all from the same ellipse, for linear combinations of the regression coefficients. In regressions employing the exponential $B_0U_1^{b_1}U_2^{b_2}\cdots U_k^{b_k}$ the degree of this homogeneous function is often of interest, and this is indicated by the sum $b_1+b_2+\cdots+b_k$. In production studies this sum indicates the degree of returns to scale;[3] in the current example, (3.2), it indicates whether a

[3] In production studies it is usually desirable to assume that the independent variables are subject to error. The treatment of this problem and a substantial bibliography are given by Tintner [10].

stock split will result in a proportional reduction in the price of the stock. The subsidiary interval for the sum b_1+b_2 in (3.2),

$$.90 \leq b_1 + b_2 \leq 1.00, \tag{3.4}$$

can be established graphically by drawing lines of the family $y=b_1+b_2$ in Figure 1 and regarding as inadmissible all those that do not meet the ellipse at any point.

The totality of all possible subsidiary intervals is, in a sense, equivalent to the joint ellipse. The combined intervals (3.3) restrict the parameter point (b_1, b_2) to the rectangle $BGEH$ of Figure 1, which includes all of the ellipse. The combination of (3.3) and (3.4) further restricts this point to the hexagon $ABCDEF$, which again includes the ellipse. If this process is repeated by combining still more intervals—say b_2-b_1 or b_2-3b_1—the resulting many-sided figure can be made to approach the ellipse as closely as desired. Hence the totality of all possible subsidiary intervals has a joint confidence level of $1-\alpha$. Accordingly, if one performs a series of experiments, he may expect that in at least $100(1-\alpha)$ per cent of them, all of the subsidiary intervals, however chosen, will cover the true parameter point.

4. SUBSIDIARY INTERVALS IN k DIMENSIONS

In a two dimensional problem like the preceding, subsidiary limits can be derived graphically with fair accuracy. But when greater accuracy is desired, or when more than two dimensions are involved, an analytic procedure is indicated. The problem in general is to set limits for linear combinations

$$Q = \sum_{i=1}^{k} h_i b_i, \tag{4.1}$$

where the constants h_i are given. If arbitrary values are assigned to Q, equation (4.1) defines a family of hyperplanes in k-dimensional parameter space. In particular there are two distinct quantities $\overline{Q}$ and $\underline{Q}$ $(\overline{Q}>\underline{Q})$ that define two planes tangent to the confidence ellipse (3.1). Hence all members of family (4.1) having the property $Q>\overline{Q}$ or $Q<\underline{Q}$ will lie completely outside the ellipse, and the subsidiary statement is, therefore, $\underline{Q}\leq Q\leq\overline{Q}$.

Finding the tangent planes and the quantities $\overline{Q}$ and Q may be facilitated by transforming the confidence ellipse (3.1) into a sphere with

its center at the origin. For this purpose a linear transformation is required[4]

$$d_i = \hat{b}_i - b_i = \sum_{j=1}^{k} c_{ij}\delta_j \tag{4.2}$$

such that

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}(\hat{b}_i - b_i)(\hat{b}_j - b_j) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}d_id_j = \sum_{i=1}^{k} \delta_i^2. \tag{4.3}$$

By means of (4.2) the confidence ellipse is transformed into the sphere

$$\frac{F_\alpha(k, n-k-1)kn\hat{\sigma}^2}{n-k-1} = \sum_{i=1}^{k} \delta_i^2 \tag{4.4}$$

and the tangent planes $\bar{Q} = \sum h_ib_i$ and $\underline{Q} = \sum h_ib_i$ are transformed into new planes tangent to this sphere. For $\bar{Q}$ (a similar relation holds for $\underline{Q}$) the new plane is

$$\sum_{i=1}^{k} h_i\hat{b}_i - \bar{Q} = \sum_{j=1}^{k} m_j\delta_j,$$

where

$$m_j = \sum_{i=1}^{k} c_{ij}h_i.$$

A solution for $\bar{Q}$ is now obtainable from

$$\frac{\bar{Q} - \sum h_i\hat{b}_i}{\sqrt{\sum m_j^2}} = \sqrt{\frac{F_\alpha(k, n-k-1)kn\hat{\sigma}^2}{n-k-1}} \tag{4.5}$$

where the left hand member is the distance between the transformed plane and the origin, and the right hand member is the radius of the transformed confidence sphere (4.4). A solution for $\underline{Q}$ can be obtained by taking a negative value for one of the square roots in (4.5). Therefore, the subsidiary interval takes the form

---

[4] Such a transformation can be found—in fact a number of different transformations can be found—if the matrix of the a_{ij}'s is of rank k (Bôcher [1, p. 134 ff.]. Moreover, the rank will be k if the X_i's are linearly independent as assumed (Wilks [12, p. 160]).

$$
\begin{aligned}
(4.6)\qquad & \sum h_i\hat{b}_i - \sqrt{\frac{F_\alpha(k, n-k-1)kn\hat{\sigma}^2\sum m_j^2}{n-k-1}} \\
& \leq \sum h_i b_i \leq \sum h_i\hat{b}_i + \sqrt{\frac{F_\alpha(k, n-k-1)kn\hat{\sigma}^2\sum m_j^2}{n-k-1}}.
\end{aligned}
$$

In the above, it can be shown (though the proof will not be given) that the quantity

$$\frac{n\hat{\sigma}^2\sum m_j^2}{n-k-1}$$

is the unbiased estimate of the variance of the linear combination $\hat{Q} = \sum h_i\hat{b}_i$. Then (4.6) can be written

$$
\begin{aligned}
(4.7)\qquad & \hat{Q} - \sqrt{kF_\alpha(k, n-k-1)\text{ Est. Var. }\hat{Q}} \\
& \leq Q \leq \hat{Q} + \sqrt{kF_\alpha(k, n-k-1)\text{ Est. Var. }\hat{Q}},
\end{aligned}
$$

which is essentially the result derived by Scheffé [9] for contrasts in the analysis of variance and by Roy and Bose [8] for the general linear case.

Of the many possible ways of deriving transformation (4.2), the one chosen for illustration here is conveniently used in conjunction with the Doolittle solution for the normal regression equations. In essence, the quadratic form on the left of (4.3) is reduced to a sum of squares by the method of Lagrange (see Bôcher [1, p. 131]). The reduction is performed by the following transformation:

$$
(4.8)\qquad
\begin{aligned}
\delta_1 &= (a_{11}d_1 + a_{12}d_2 + a_{13}d_3 + \cdots + a_{1k}d_k)/\sqrt{a_{11}} \\
\delta_2 &= \qquad\qquad (a_{22}'d_2 + a_{23}'d_3 + \cdots + a_{2k}'d_k)/\sqrt{a_{22}'} \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\delta_k &= \qquad\qquad\qquad\qquad a_{kk}^{(k-1)}d_k/\sqrt{a_{kk}^{(k-1)}}
\end{aligned}
$$

in which

$$
\begin{aligned}
a_{ij}' &= a_{ij} - a_{i1}a_{1j}/a_{11} \\
a_{ij}'' &= a_{ij}' - a_{i2}'a_{2j}'/a_{22}' \\
& \cdots\cdots\cdots\cdots
\end{aligned}
$$

Thus, all of the entries in (4.8) arise in the regular reduction process of the forward Doolittle solution. Finally, the required transformation (4.2) is obtained by inverting (4.8), which is easy.

As a numerical example, Table 2 shows the standard forward Doolittle solution for the regression coefficients of (2.1) and (2.2). The standard form is commonly found in texts (see, for example, Croxton and Cowden [2, pp. 716–20]) and is therefore suitable for illustration. For practical computation, the abbreviated Doolittle is probably superior (see Dwyer [3, pp. 107–12]). Transformation (4.8) would be derived from the italicized quantities in Table 2 as follows

$$\begin{aligned}
\delta_1 &= (3.2554d_1 + 3.5128d_2 + 3.4260d_3)/1.8043 \\
\delta_2 &= \qquad\qquad\quad (3.4688d_2 - .1432d_3)/1.8625 \\
\delta_3 &= \qquad\qquad\qquad\qquad\qquad .0870d_3/.2950.
\end{aligned}$$

This is easily inverted by solving for the d_i's in terms of the δ_i's, thus

$$\begin{aligned}
d_1 &= .5542\delta_1 - .5794\delta_2 - 3.7189\delta_3 \\
d_2 &= \qquad\qquad\quad .5369\delta_2 + .1400\delta_3 \qquad (4.2') \\
d_3 &= \qquad\qquad\qquad\qquad\quad 3.3903\delta_3.
\end{aligned}$$

Now suppose that a subsidiary statement is desired for the sum $b_1+b_2+b_3$ in (2.1), whose maximum likelihood estimate is $-.04$. The same subsidiary statement is applicable, of course, to the exponent of C in (2.3) and of S in (2.4). By means of (4.2′) the sum is transformed thus

$$\begin{aligned}
\hat{b}_1 + \hat{b}_2 + \hat{b}_3 - \bar{Q} &= d_1 + d_2 + d_3 \\
&= \sum m_j\delta_j \\
&= .5542\delta_1 - .0425\delta_2 - .1886\delta_3.
\end{aligned}$$

From the above, one calculates the quantity

$$\sum m_j^2 = .3445.$$

This is substituted in (4.6) along with $k=3$, $n=17$, $\hat{\sigma}^2=.0006863$, and $F_{.05}(3,\ 13)=3.410$, and the resulting interval is $-.096 \leqq b_1+b_2+b_3 \leqq .016$.

5. THE JOINT CONFIDENCE ELLIPSE VERSUS CONVENTIONAL CONFIDENCE INTERVALS

Since Scheffé [9] has discussed at some length the relation between the confidence ellipse and the conventional form of confidence interval for the analysis of variance, only a brief extension to regression problems is needed here. In Figure 2, which again refers to (3.2), the con-

ventional 95 per cent limits, already given in (1.3), are shown superimposed on the ellipse of Figure 1. The hexagon $A'B'C'D'E'F'$ formed by the intersection of the three conventional bands is similar and similarly situated to the circumscribed hexagon $ABCDEF$ in Figure 1, and one may surmise that the smaller hexagon encloses an ellipse similar to the 95 per cent ellipse. Geometrically, then, the conventional

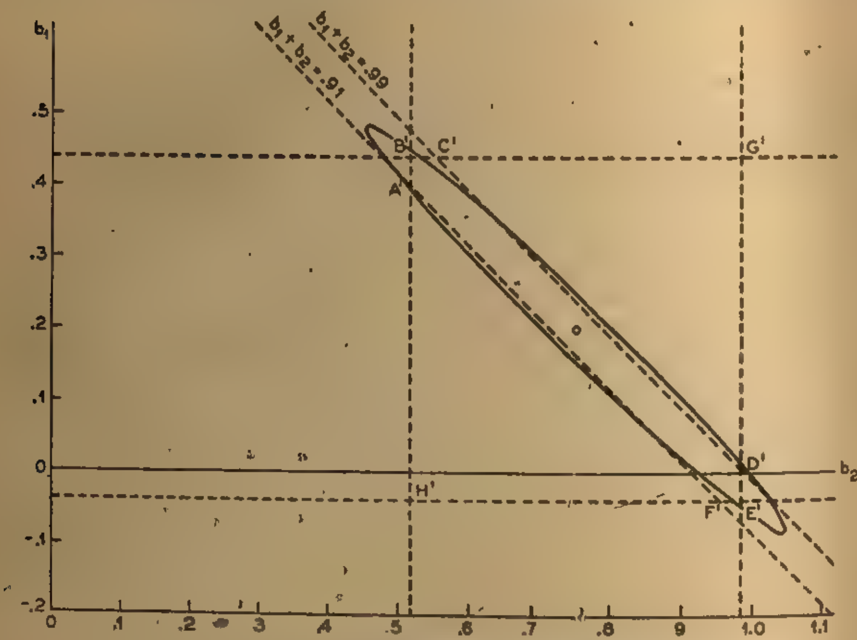


FIG. 2. 95 per cent joint confidence ellipse and conventional 95 per cent confidence intervals for (3.2).

intervals of Figure 2 bear the same relation to their inscribed ellipse that the subsidiary intervals bear to the 95 per cent ellipse; that is the smaller elipse is equivalent to the totality of all possible conventional intervals and provides a lower limit for the joint confidence level of a set of statements like (1.3).

To determine the confidence level of the ellipse inscribed within $A'B'C'D'E'F'$, it is convenient to rewrite (1.2)

$$\hat{b}_p - t_\alpha\sqrt{\text{Est. var. } \hat{b}_p} \leq b_p \leqq \hat{b}_p + t_\alpha\sqrt{\text{Est. var. } \hat{b}_p}.$$

Then, on comparison with (4.7), it is apparent that

$$t_{\alpha''} = \sqrt{kF_{\alpha'}(k, n - k - 1)} \tag{5.1}$$

or

$$F_{\alpha''}(1, n - k - 1) = kF_{\alpha'}(k, n - k - 1),$$

where $1-\alpha''$ is the confidence level of the conventional interval and $1-\alpha'$ is the level of the inscribed ellipse. The problem of finding α' given α'', or vice versa, is conveniently solved by means of Pearson's *Tables of the Incomplete Beta-Function* [7]. In the example under discussion, $k=2$ and $t_{.05}=2.145$ for fourteen degrees of freedom; then (5.1) indicates $F\ (2, 14)=2.3005$. By means of the transformation described by Pearson [7, p. xlvii] or Mood [5, p. 206], approximately .86 for $1-\alpha'$ is obtained from the beta-function table. Thus the probability that (1.3) is correct is bounded by .86 and .95. Table 3 presents values of $1-\alpha'$ corresponding to the conventional .95 and .99 levels of $1-\alpha''$ for several other combinations of k and $n-k-1$. For large samples and special situations where the variance is known, the limiting value of $1-\alpha'$ for $F_{\alpha'}(k, \infty)$ is obtained from Pearson's *Tables of the Incomplete Gamma-Function* [6].

6. INCLUSION OF b_0 IN THE CONFIDENCE ELLIPSE

In some applications it may be desirable to extend the confidence ellipse so as to cover the constant term b_0. Define an arbitrary independent variable $X_0=1$. Then, b_0 is merely the partial regression coefficient of Y on X_0. Let

$$\sum_{\xi=1}^{n} X_{i\xi}X_{j\xi} = A_{ij}.$$

In particular, $A_{00}=\sum X_{0\xi}{}^2=n$ and $A_{0j}=\sum X_{0\xi}X_{j\xi}=\sum A_{j\xi}$. Then, as shown by Wilks [12, Sect. 8.3] and Mood [5, Sect. 13.5], the joint confidence ellipsoid is defined by

$$F_\alpha(k+1, n-k-1) \doteq \frac{(n-k-1)\sum_{i=0}^{k}\sum_{j=0}^{k} A_{ij}(\hat{b}_i - b_i)(\hat{b}_j - b_j)}{(k+1)n\hat{\sigma}^2}.$$

This, like (3.1) may be transformed into a sphere by the method of Section 4.

TABLE 3

LOWER BOUND OF PROBABILITY THAT ALL CONVENTIONAL CONFIDENCE INTERVALS WILL BE CORRECT FOR k REGRESSION COEFFICIENTS IN SAMPLES OF n

Conventional Confidence Level	Degrees of Freedom $n-k-l$	Number of Independent Variables k						
		2	3	4	5	6	8	10
.95	5	.88	.79	.70	.62	.53	.38	.27
	7	.87	.78	.67	.57	.47	.31	.19
	10	.87	.76	.65	.53	.43	.26	.14
	15	.86	.75	.62	.50	.39	.21	.11
	20	.86	.74	.61	.48	.37	.19	.09
	30	.86	.74	.60	.46	.34	.17	.07
	60	.86	.73	.59	.45	.32	.15	.06
	∞	.85	.72	.57	.43	.30	.13	.05
.99	5	.97	.95	.92	.89	.85	.77	.69
	7	.97	.94	.91	.86	.81	.71	.59
	10	.97	.94	.89	.84	.78	.64	.50
	15	.97	.93	.88	.81	.74	.58	.42
	20	.97	.93	.87	.80	.72	.54	.38
	30	.97	.92	.86	.78	.70	.50	.33
	60	.96	.92	.85	.77	.67	.47	.29
	∞	.96	.92	.84	.75	.64	.42	.24

CONCLUSION

In the simon pure application of the conventional approach, where a single statement is specified, it is possible to establish an interval with a definite probability $1-\alpha$. The same is true for certain special problems involving multiple statements—such as Tukey's problem of contrasting means. But in general, we must conclude that establishing a single probability for a set of multiple statements is either impracticable or impossible. Instead, there will be two bounds, $1-\alpha'$ and $1-\alpha''$ between which lies the probability of being right. How, then, are these gounds to be chosen?

The common procedure to date, of establishing the upper bound without regard to the lower, can hardly be justified. To take an extreme example from Table 3, what does it mean if we establish intervals for a 10-variable regression and all we can say is that the probabil-

ity of being right lies somewhere between .05 and .95? The converse procedure, of establishing the lower bound without regard to the upper, as recommended here, will be criticized, no doubt, as unduly conservative. But whether this criticism is justified or not will depend on the number of statements contemplated and the degree of flexibility required in choosing them. In the bank stock example, where a multiplicity of statements is indicated and the choice of statement must be deferred until after the experiment in order to avoid missing important findings, there is every reason to believe that the true probability $1-\alpha$ is much closer to its lower bound than to its upper. Here the loss of efficiency in using the conservative approach cannot be very great.

There remains, of course, the perplexing problem of establishing bounds when a limited number of statements are contemplated but no convenient computation procedure is available for ascertaining $1-\alpha$. Then, a possible solution is to select bounds that straddle the desired probability. For small values of k, say less than four, the bounds are not too far apart, and one might find examples for which $1-\alpha'=.90$ and $1-\alpha''=.99$ approximately. Perhaps this provides a working approximation for the .95 level on the hope that $1-\alpha$ lies about midway between its bounds.

ACKNOWLEDGMENTS

I wish to thank J. Arthur Greenwood and W. Braddock Hickman for a great deal of inspiration and advice. However, I must absolve both of these gentlemen from any responsibility—particularly since one of them has reservations.

REFERENCES

[1] Bôcher, Maxime, *Introduction to Higher Algebra*, New York, The Macmillan Company, 1936.

[2] Croxton, F. E. and Cowden, D. J., *Applied General Statistics*, New York, Prentice-Hall, 1939.

[3] Dwyer, P. S., *Linear Computations*, New York, John Wiley and Sons, 1951.

[4] Haavelmo, Trygve, "Methods of Measuring the Marginal Propensity to Consume," *Journal of the American Statistical Association*, 42 (1947), 105–22.

[5] Mood, Alexander M., *Introduction to the Theory of Statistics*, New York, McGraw-Hill Book Company, 1950. See especially Chapter 13.

[6] Pearson, Karl, editor, *Tables of the Incomplete Gamma-Function*, London, The *Biometrika* Office, reissue, 1946.

[7] Pearson, Karl, editor, *Tables of the Incomplete Beta-Function*, London, The *Biometrika* Office, 1934.

[8] Roy, S. N. and Bose, R. C., "Simultaneous Confidence Interval Estimation," *Annals of Mathematical Statistics*, 24 (1953), 513–36.

[9] Scheffé, Henry, "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika*, 40 (1953), 87–104.

[10] Tintner, Gerhard, "A Test for Linear Relations between Weighted Regression Coefficients," *Journal of the Royal Statistical Society*, Series B, 12 (1950), 273–77.

[11] Tukey, John W., "Allowances for Various Types of Error Rates," unpublished invited address presented before a joint meeting of the Institute of Mathematical Statistics and the Biometrics Society (ENAR) at Blacksburg, Va., March 19, 1952. For discussion of Tukey's method see [8] and [9].

[12] Wilks, S. S., *Mathematical Statistics*, Princeton University Press, Princeton, N. J., 1946. See especially Chapter 8.

ASYMPTOTIC RELATIVE EFFICIENCIES OF DISTRIBUTION-FREE TESTS OF RANDOMNESS AGAINST NORMAL ALTERNATIVES

ALAN STUART
London School of Economics

1. THE MEASURE OF EFFICIENCY

SEVERAL writers, notably Hotelling and Pabst [5], have explicitly assumed that the relative efficiency of two test statistics is to be measured by their estimating efficiencies. While this seems reasonable, it is by no means obvious, since if the two tests are consistent, the ratio of their powers against any *fixed* alternative hypothesis must tend to unity with increasing sample size n, and it may easily be shown that for any n, the less efficient estimator may provide a more powerful test (Sundrum [14]).

Pitman [11] has proposed a measure of the asymptotic relative efficiency of consistent tests. Given that the two statistics, t_1 and t_2, have normal limit distributions with variances of order n^{-1}, and that certain general regularity conditions are satisfied, he considered a limiting process in which the alternative hypothesis H_1 differs from the null hypothesis H_0 by a quantity of order $n^{-1/2}$, so that as n increases, H_1 tends to H_0. Under these conditions, he showed that the reciprocal of the ratio of sample sizes required to attain equal power against the same alternative was, in the limit,

$$\frac{\left\{\left[\frac{\partial}{\partial\theta}E(t_1)\right]_{\theta=\theta_0}\right\}^2}{V(t_1\mid\theta=\theta_0)}\cdot\frac{V(t_2\mid\theta=\theta_0)}{\left\{\left[\frac{\partial}{\partial\theta}E(t_2)\right]_{\theta=\theta_0}\right\}^2},\tag{1}$$

where θ is the parameter whose value distinguishes H_0 from H_1, and E and V denote mean value and variance respectively.

Some such limiting process is necessary if we require a single measure of the relative efficiency of two tests, but it is not altogether surprising that Pitman's result is equivalent to the use of estimating efficiency as a criterion. With t_1 and t_2 as our (consistent) test statistics, let

$$T_1=f(t_1),\quad T_2=g(t_2)$$

be transformations of them which are consistent *estimators* of the underlying parameter θ. For large samples,

$$V(T_1) = \left(\frac{\partial f}{\partial t_1}\right)^2 V(t_1).$$

If $T_1=\theta\{1+0(n^{-\delta})\}$ and $t_1=E(t_1)\{1+0(n^{-\epsilon})\}$ where δ, ϵ are positive constants, we may write,[1] to our order of approximation,

$$\frac{\partial f}{\partial t_1} = \frac{\partial f}{\partial \theta}\bigg/\frac{\partial t_1}{\partial \theta} = 1\bigg/\frac{\partial E(t_1)}{\partial \theta}.$$

Thus

$$V(T_1) = V(t_1)\bigg/\left\{\frac{\partial E(t_1)}{\partial \theta}\right\}^2.$$

Similarly

$$V(T_2) = V(t_2)\bigg/\left\{\frac{\partial E(t_2)}{\partial \theta}\right\}^2.$$

Thus

$$\frac{V(T_2)}{V(T_1)} = \frac{(E_1')^2}{v_1}\cdot\frac{V_2}{(E_2')^2} \tag{2}$$

in a shortened notation.

At θ_0 the right side of (2) is equivalent to Pitman's measure (1), and since the left side of (2) is the ordinary estimating efficiency of the transformed statistics on H_0, it follows that Pitman's measure simply reproduces the estimating properties of the appropriate transformations of the test-statistics being considered. As a simple example, the sign test for the median reproduces the efficiency property of the sample median as an estimator of the population mean. (The result is $2/\pi$ for normal populations.)

Thus Pitman's result on power may be regarded as a justification of the procedure of using estimating efficiency as a test criterion.

2. TESTS OF RANDOMNESS AGAINST NORMAL REGRESSION ALTERNATIVES

Using Pitman's measure, we may investigate tests of randomness for the standardised normal regression model

$$y_i = \alpha + \beta X_i + \epsilon_i,$$

---

[1] Mr. J. Durbin has considerably simplified this derivation.

where ϵ is a normal vector with $E(\epsilon)=0$ and $V(\epsilon)=1$. If the values of X are spaced at equal intervals, no generality is lost in replacing them by the numbers 1 to n.

With this model, H_0 is the hypothesis that $\beta=0$, so that β is the underlying parameter.

The standard test in this situation is based on the sample statistic

$$b = \sum_i (y_i - \bar{y})(X_i - \bar{X}) \Big/ \left\{ \sum_i (X_i - \bar{X})^2 \right\},$$

which is exactly normally distributed with mean β and variance $1/\{\sum_i(X_i-\bar{X})^2\}$. Since we have replaced the X_i by the natural numbers, $\sum_i(X_i-\bar{X})^2=n(n^2-1)/12$ and, for the statistic b, we have

$$\frac{(E')^2}{V} \sim \frac{n^3}{12}. \tag{3}$$

3. PRELIMINARY RESULTS

For application in succeeding sections, we require a few preliminary results for the normal regression model.

Define

$$H_{ij}\begin{cases} = 1 \text{ if } y_i>y_j, \\ = 0 \text{ if } y_i<y_j, \end{cases}$$

and

$$H_{ii} = 1.$$

(a) Now $E(H_{ij})=\text{Prob}\{H_{ij}=1\}$, and since (y_i-y_j) is a normal variate with mean $\beta(i-j)$ and variance 2, this is

$$= \int_0^\infty N\{\beta(i-j), 2\}$$

$$= \int_{-\beta(i-j)/2}^\infty N(0, 1),$$

in an obvious notation.

Thus

$$\left[\frac{\partial}{\partial\beta} E(H_{ij})\right]_{\beta=0} = \frac{(i-j)}{\sqrt{2}}\cdot\frac{1}{\sqrt{2\pi}} = \frac{(i-j)}{2\sqrt{\pi}}. \tag{4}$$

(b) Also $E(H_{ij}H_{kl})=E(H_{ij})E(H_{kl})$.

So that

$$\frac{\partial}{\partial\beta}E(H_{ij}H_{kl}) = E(H_{ij})\left\{\frac{\partial}{\partial\beta}E(H_{kl})\right\} + \left\{\frac{\partial}{\partial\beta}E(H_{ij})\right\}E(H_{kl})$$

and, from (4),

$$\begin{aligned}\left[\frac{\partial}{\partial\beta}E(H_{ij}H_{kl})\right]_{\beta=0} &= \frac{1}{2}\left(\frac{k-l}{2\sqrt{\pi}} + \frac{i-j}{2\sqrt{\pi}}\right) \\ &= \frac{1}{4\sqrt{\pi}}\{(i+k)-(j+l)\}.\end{aligned} \tag{5}$$

(c) Since

$$\frac{d}{dx}\phi(ax, bx) = a\frac{\partial\phi}{\partial(ax)} + b\frac{\partial\phi}{\partial(bx)},$$

we have

$$\frac{d}{dx}\int_{-\infty}^{ax}\int_{-\infty}^{bx} f(u,v)du dv = a\int_{-\infty}^{bx} f(ax,v)dv + b\int_{-\infty}^{ax} f(u,bx)du, \tag{6}$$

and similarly for variables at the lower limits of integration.

Now $(y_i - y_j)$ and $(y_j - y_k)$ are jointly normally distributed with correlation $-\frac{1}{2}$. Thus

$$\begin{aligned}E(H_{ij}H_{jk}) &= \text{Prob }(H_{ij}=1, H_{jk}=1) \\ &= \int_{-\beta(i-j)/\sqrt{2}}^{\infty}\int_{-\beta(j-k)/\sqrt{2}}^{\infty} N\begin{pmatrix}0, & 1 & -\frac{1}{2} \\ 0, & 1 & \end{pmatrix},\end{aligned}$$

in an obvious notation, so that, applying (6), appropriately modified,

$$\begin{aligned}\left[\frac{\partial}{\partial\beta}E(H_{ij}H_{jk})\right]_{\beta=0} &= \left(\frac{i-j}{\sqrt{2}} + \frac{j-k}{\sqrt{2}}\right)\frac{1}{\pi\sqrt{3}}\cdot\frac{1}{2}\cdot\frac{\sqrt{3}}{2}\cdot\sqrt{2\pi} \\ &= \frac{i-k}{4\sqrt{\pi}}.\end{aligned} \tag{7}$$

Similarly, we see that

$$\left[\frac{\partial}{\partial\beta}E(H_{ik}H_{jk})\right]_{\beta=0} = \frac{i+j-2k}{4\sqrt{\pi}}. \tag{8}$$

Hence, summarising results (4)–(8) and remembering that $H_{ii}=1$ identically, we have

$$\left.\begin{aligned}\left[\frac{\partial}{\partial\beta}E(H_{ij}H_{kl})\right]_{\beta=0} &= \frac{i+k-\overline{j+l}}{4\sqrt{\pi}} \text{ if } j\neq i \text{ and } l\neq k\\ \left[\frac{\partial}{\partial\beta}E(H_{ii}H_{kl})\right]_{\beta=0} &= \frac{k-l}{2\sqrt{\pi}} \text{ for all } k, l\end{aligned}\right\}. \tag{9}$$

(d) Finally, we consider the probability that the product $(y_i-y_{i-1})(y_{i-1}-y_{i-2})$ is negative. This requires that one of the brackets is positive and the other negative, and this has probability

$$E = 2\int_{-\infty}^{-\beta/\sqrt{2}}\int_{-\beta/\sqrt{2}}^{\infty} N\begin{pmatrix}0, & 1 & \\ 0, & 1 & -\tfrac{1}{2}\end{pmatrix},$$

and, applying (6), suitably modified, we find that the terms cancel and

$$\left[\frac{\partial E}{\partial\beta}\right]_{\beta=0} = 0. \tag{10}$$

We now proceed to calculate $(E')^2/V$ for five well-known distribution-free tests of randomness, where the null hypothesis is that all the observations have come from the same continuous population. In the normal regression model, this is equivalent to testing $H_0:\beta=0$.

4. THE DIFFERENCE-SIGN TEST

The test criterion, proposed by Moore and Wallis [8], is simply the number of positive first differences in the series. Stuart [12] has made explicit power calculations against normal regression alternatives. The test statistic is

$$D = \sum_{i=2}^{n} H_{i,i-1}.$$

By (4),

$$E'(D) = \frac{(n-1)}{2\sqrt{\pi}}.$$

Also, as is easily shown (Stuart [12]),

$$V(D) = \frac{1}{12}(n+1).$$

So for D,

$$\frac{(E')^2}{V} \sim 3n/\pi. \qquad (11)$$

5. KENDALL'S RANK CORRELATION TEST

Mann [7] suggested the use of Kendall's rank correlation coefficient, t, for testing randomness. We shall consider the quantity Q related to t by

$$t = 1 - \frac{4Q}{n(n-1)}.$$

Q may be defined by

$$Q = \sum_{i<j}^{n} H_{ij},$$

i.e. we take all possible $\frac{1}{2}n(n-1)$ comparisons between pairs of observations, and score unity whenever an observation exceeds a later observation.

From (4),

$$E'(Q) = -\sum_{i<j}^{n} (j-i)/(2\sqrt{\pi})$$

$$= -\frac{n(n^2-1)}{12\sqrt{\pi}}.$$

Also (see, e.g., Kendall [6])

$$V(Q) \sim n^3/36$$

so that for Q,

$$\frac{(E')^2}{V} \sim n^3/(4\pi). \qquad (12)$$

6. SPEARMAN'S RANK CORRELATION TEST

Spearman's rank correlation coefficient, r_s, can clearly be used wherever t can, and has been considered by Daniels [2] as a test against trend. We shall consider the quantity V related to r_s by

$$r_s = 1 - \frac{12V}{n(n^2-1)}.$$

It may easily be shown (see Durbin and Stuart [3]) that V may be defined analogously to Q by

$$V = \sum_{i<j}^{n} (j - i)H_{ij},$$

i.e. V is a weighted sum of the H-scores, the weights being the distance separating the observations contributing the score. Now

$$E(V) = \sum_{i<j} (j - i)E\{H_{ij}\}$$

and, by (4),

$$E'(V) = -\frac{1}{2\sqrt{\pi}} \sum_{i<j} (j - i)^2$$
$$= -\frac{n^2(n^2 - 1)}{24\sqrt{\pi}};$$

Also (see Kendall [6]) it is easily shown that the variance of V is asymptotically $n^5/144$.

Thus

$$\frac{(E')^2}{v} \sim n^3/(4\pi), \tag{13}$$

just as for Kendall's test.

7. THE TURNING POINT TEST

Another test, proposed by Wallis and Moore [16], consists in counting the number of runs up and down in the series or, equivalently, the number of peaks and troughs in the series. This statistic was considered earlier from another point of view by Bilham [1]. We score

$$T_i = \begin{cases} +1 & \text{if} \quad (y_i - y_{i-1})(y_{i-1} - y_{i-2}) < 0 \\ 0 & \text{otherwise,} \end{cases}$$

and the sum of the $(n-2)\,T_i$ is our statistic T. Now we have seen in equation (10) that

$$\left[\frac{\partial E(T_i)}{\partial \beta}\right]_{\beta=0} = 0.$$

Thus, for T also,

$$E' = 0 \tag{14}$$

and the relative efficiency of the test is zero.

8. THE RANK SERIAL CORRELATION TEST

Since $H_{ii}=1$, the *rank* of y_i among the n y's is

$$r_i = \sum_{j=1}^{n} H_{ij}.$$

Consider

$$\begin{aligned} r_i r_k &= \sum_{j=1}^{n} \sum_{l=1}^{n} H_{ij} H_{kl} \\ &= \sum_{j\neq i} \sum_{l\neq k} H_{ij} H_{kl} + \sum_{l\neq k} H_{ii} H_{kl} \\ &\quad + \sum_{j\neq i} H_{ij} H_{kk} + H_{ii} H_{kk}. \end{aligned}$$

From (9) we obtain immediately

$$\begin{aligned} \left[\frac{\partial}{\partial\beta} E(r_i r_k)\right]_{\beta=0} &= \frac{1}{4\sqrt{\pi}} \left\{ \sum_{j\neq i} \sum_{l\neq k} (\overline{i+k} - \overline{j+l}) \right. \\ &\quad \left. + 2\sum_{l\neq k} (k - l) + 2\sum_{j\neq i} (i - j) \right\} \\ &= \frac{n(n+1)}{4\sqrt{\pi}} (i + k - \overline{n+1}). \end{aligned} \tag{15}$$

Now the rank serial correlation coefficient of lag s is, neglecting constants,

$$W = \sum_{i=1}^{n-s} r_i r_{i+s}, \tag{16}$$

when a non-circular definition is used, or

$$W' = W + \sum_{i=1}^{s} r_{n-s+i} r_i, \tag{17}$$

when a circular definition is used. These statistics are special cases, using ranks of the serial correlation coefficient proposed as a distribution-free test against trend by Wald and Wolfowitz [15].

If in (15), we put $k=i+s$, the non-constant factor becomes $(2i-\overline{n-s+1})$, and since

$$\sum_{i=1}^{s-1} (2i - \overline{n - s + 1}) = 0,$$

we see from (16) that

$$\left[\frac{\partial}{\partial\beta} E(W)\right]_{\beta=0} = 0. \tag{18}$$

Similarly, if we put $k = n - s + i$ in (15), the non-constant factor becomes $(2i - \overline{s+1})$, and since

$$\sum_{i=1}^{s} (2i - \overline{s + 1}) = 0,$$

we see from (17) and (18) that

$$\left[\frac{\partial}{\partial\beta} E(W')\right]_{\beta=0} = 0. \tag{19}$$

Thus, for any lag, the circular or non-circular rank serial correlation coefficient has relative efficiency zero as a test against normal regression.

9. COMPARISONS AND CONCLUSIONS

Collecting our results (3), (11), (12), (13), (14), (18) and (19), we obtain, using (1), the following table for the asymptotic relative efficiencies of the tests:

Test		Asymptotic Relative Efficiency: Compared to b	Asymptotic Relative Efficiency: Compared to D
Regression coefficient	(b)	1	—
Spearman's test	(V)	$3/\pi = .95$	
Kendall's test	(Q)	$3/\pi = .95$	
Difference-sign test	(D)	0	1
Turning point test	(T)	0	0
Rank serial correlation test	(W)	0	0

The rank correlation tests are highly and equally efficient, agreeing with the results of Daniels [2], who considers a more general model. Our value $3/\pi$ is not identical with the value $9/\pi^2$ established for their efficiency as tests of bivariate independence by Hotelling and Pabst

[5] and by Moran [9]: here we have been dealing with an essentially univariate problem, and since the null hypothesis is one of independence, it is not surprising that we get for our efficiency the square root of the bivariate efficiency. (An "explanation" of this result is given by Stuart [13].) When we take into consideration the computing time for each statistic, Spearman's test is to be preferred, the formula

$$V = \frac{1}{2} \sum_{i=1}^{n} (r_i - i)^2$$

giving it a clear advantage over Kendall's test, especially for large n.

As will be seen from the table, all three other tests have asymptotic relative efficiencies of zero, although, as the last column shows, D is to be preferred to the other two, which have zero relative efficiency for any value of n. (The result for W has previously been obtained by Noether [10].) Care must be taken in interpreting these results, for, since the tests are all consistent against this alternative hypothesis, we can always make the power of any of them as close to 1 as we please by increasing sample size indefinitely. The situation is analogous to (and, as shown in section 1 above, a reflection of) that arising with a consistent, but highly inefficient, estimator. The relative efficiencies which we have calculated are local properties, in the sense that they are restricted to neighbourhoods of the null hypothesis which are small enough, when sample size is taken into account, to keep the powers of the tests bounded away from unity. Reference should be made here to a sampling experiment reported by Foster and Stuart [4] which closely bears out the results of the table above.

While D is very much simpler to compute than any of the other statistics, this fact does not close the gap between it and V. For (11) and (13) show that V has an efficiency advantage of order n^2, while if computing time is proportional to the number of comparisons to be made between observations, the advantage to D with $(n-1)$ comparisons, as against $\frac{1}{2}n(n-1)$ for V, is only of order n.

10. SUMMARY

It is shown that against normal regression alternatives, the two rank correlation tests are to be preferred to three other distribution-free tests, these being the difference-sign test, the rank serial correlation coefficient test and the turning point test. Further, on computational grounds alone, Spearman's rank correlation test is to be preferred to Kendall's test. These results do not, of course, apply to other alternative hypotheses, such as the presence of serial correlation.

ACKNOWLEDGMENT

I should like to express my thanks to Dr. H. R. van der Vaart of Leiden University for a close critical reading of the first draft of this paper, which led to the removal of a number of errors.

REFERENCES

[1] Bilham, E. G., "Correlation coefficients," *Quarterly Journal of the Royal Meteorological Society*, 52 (1926), 172.

[2] Daniels, H. E., "Rank correlation and population models," *Journal of the Royal Statistical Society, Series B (Methodological)*, 12 (1950), 171–81.

[3] Durbin, J., and Stuart, A., "Inversions and rank correlation coefficients," *Journal of the Royal Statistical Society, Series B (Methodological)*, 13 (1951), 303–9.

[4] Foster, F. G., and Stuart, A., "Distribution-free tests in time-series based on the breaking of records," *Journal of the Royal Statistical Society, Series B (Methodological)*, 16 (1954).

[5] Hotelling, Harold, and Pabst, Margaret Richards, "Rank correlation and tests of significance involving no assumption of normality," *Annals of Mathematical Statistics*, 7 (1936), 29–43.

[6] Kendall, M. G., *Rank Correlation Methods*, London, Griffin, 1948.

[7] Mann, Henry B., "Non-parametric tests against trend," *Econometrica*, 13 (1945), 245–59.

[8] Moore, Geoffrey H., and Wallis, W. Allen, "Time series significance tests based on signs of differences," *Journal of the American Statistical Association*, 38 (1943), 153–64.

[9] Moran, P. A. P., "Partial and multiple rank correlation," *Biometrika*, 38 (1951), 26–32.

[10] Noether, Gottfried E., "Asymptotic properties of the Wald-Wolfowitz test of randomness," *Annals of Mathematical Statistics*, 21 (1950), 231–46.

[11] Pitman, E. J. G., "Lecture notes on non-parametric inference," (Univ. of N. Carolina, mimeographed, 1948).

[12] Stuart, Alan, "The power of two difference-sign tests," *Journal of the American Statistical Association*, 47 (1952), 416–24.

[13] Stuart, Alan, "The correlation between variate-values and ranks in samples from a continuous distribution," *British Journal of Statistical Psychology*, (to be published).

[14] Sundrum, R. M., "Theory and application of distribution-free methods," (*Unpublished Ph.D. thesis, University of London, 1953*).

[15] Wald, A., and Wolfowitz, J., "An exact test for randomness in the non-parametric case based on serial correlation," *Annals of Mathematical Statistics*, 14 (1943), 378–88.

[16] Wallis, W. Allen, and Moore, Geoffrey H., "A significance test for time-series analysis," *Journal of the American Statistical Association*, 36 (1941), 401–9.

ESTIMATION OF THE POISSON PARAMETER FROM TRUNCATED SAMPLES AND FROM CENSORED SAMPLES*

A. C. Cohen, Jr.
University of Georgia

Maximum likelihood estimators of the Poisson parameter applicable to both truncated and censored samples are derived in this paper. Singly and doubly truncated samples as well as singly and doubly censored samples are considered. The estimators obtained are presented in simple algebraic forms and their application to practical problems with the aid of standard Poisson tables is illustrated with numerical examples. Asymptotic variances of estimates for the different cases considered are obtained from second derivatives of the likelihood functions and are simplified to forms which permit ready evaluation.

1. INTRODUCTION

The Poisson distribution is an appropriate mathematical model for studying such diverse classes of discrete data as haemocytometer counts of blood cells per square, the number of noxious weed seed per unit of field seed, and the number of defects per unit of a manufactured product. It is thus of interest to the biologist, the agronomist, and the quality control engineer as well as to research workers in various other fields of scientific endeavor. When sample observation is permitted over the full range of the complete distribution, the estimation problem is quite simple. In that case, the maximum likelihood estimate of the population parameter is the sample mean. When the sample is truncated or otherwise restricted, as for example when the number of zero observations is unknown or when observations of higher counts are pooled, the estimation problem increases in complexity. Various aspects of estimation involving singly truncated and singly censored Poisson samples with known terminals have been considered by Tippett [7], Bliss [1], Rider [5], David and Johnson [2], and by Moore [4]. According to terminology which has recently come into popular usage, truncated samples are understood to be those from which the number of observations eliminated by the restricting process is unknown. Censored samples are those in which the total number of sample specimens

* Preliminary report presented before American Mathematical Society, Auburn, Alabama, November 23, 1951. Abstract published in *Bulletin of the American Mathematical Society*, 58 (1952), 60
Sponsored in part by the Office of Ordnance Research, U.S. Army, under contract DA-01-009-ORD-288.

is known, but measurements on some of this number is lacking. Censored samples may thus be regarded as truncated samples having a known number of unmeasured (missing) observations. In this paper and in the references cited, interest lies only with samples truncated or censored such that all observations above or below specified terminals are either eliminated or unmeasured; that is, with samples in which the restriction applies only to the tails of the sample. The classification of single or double indicates whether one or both tails have been restricted. Tippett obtained the maximum likelihood estimator for a sample that is singly censored on the right, but left his results in a somewhat unwieldy form for practical application when more than four of the individual frequency classes are available. For four or less frequency classes, he provided nomograms to aid in computing the required estimates. Bliss developed an approximation to Tippett's estimator and provided two tables necessary for applying his procedure. Moore was also concerned with this case and developed an estimator based on sample moment functions. Rider developed an estimator based on moment functions for the case of samples singly truncated on the left and also considered maximum likelihood estimation for the same case. David and Johnson were likewise concerned with maximum likelihood estimation in this latter case when the zero frequency class is the only one missing. The present paper is more general and of greater extent than the references cited above. It is concerned with maximum likelihood estimation from singly and doubly truncated samples as well as from singly and doubly censored samples, all with known terminals. Estimators derived here are expressed in simple algebraic forms for easy application to practical problems.

2. THE POISSON DISTRIBUTION

The complete Poisson distribution function may be expressed as

$$f(x, m) = \frac{e^{-m}m^x}{x!}, \qquad x = 0, 1, 2, \cdots, \tag{1}$$

where $f(x, m)$ is thus the probability of observing exactly x occurrences of the event studied. The cumulative probability that c or more occurrences will be observed, may be written as

$$P(c, m) = \sum_{x=c}^{\infty} \frac{e^{-m}m^x}{x!}. \tag{2}$$

As stated in the introduction, the maximum likelihood estimate of m based on a complete (unrestricted) random sample is given as

$$\hat{m} = \bar{x} = \sum_{1}^{n} x_i/n, \tag{3}$$

where n is the total number of sample observations. In what follows, corresponding estimators are derived for various types of truncated and censored samples. Where no confusion will likely result, the notation is simplified by writing $f(x)$ and $P(c)$ in place of the longer $f(x, m)$ and $P(c, m)$.

3. TRUNCATED SAMPLES—NUMBER UNMEASURED (MISSING) OBSERVATIONS UNKNOWN

Doubly truncated. The probability function of a Poisson distribution truncated on the left at $x=c$ and on the right at $x=d$, may be written as

$$\begin{cases} f(x) = 0, & x < c, \\ f(x) = [P(c) - P(d+1)]^{-1}\dfrac{e^{-m}m^x}{x!}\cdot & c \leqq x \leqq d, \\ f(x) = 0, & x > d. \end{cases} \tag{4}$$

The truncated distribution (4) is thus normalized so that

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=c}^{d} f(x) = 1. \tag{5}$$

The likelihood function of a random sample of n observations from a population distributed according to (4) may be written as

$$P(x_1, x_2, \cdots, x_n) = [P(c) - P(d+1)]^{-n}e^{-nm}m^{\sum_1^n x_i}\left[\prod_1^n x_i!\right]^{-1}. \tag{6}$$

We obtain this same likelihood function when we consider the population as being complete with frequency function (1) and consider the sample as being truncated with the restriction that sample observation can include neither count nor measurement beyond the truncation points. In general, the writer prefers to view the truncation or other restrictions as being imposed on the sample rather than on the population.

Taking logarithms of (6) and writing $n\bar{x}$ in place of $\sum_1^n x_i$, where $\bar{x}$ is thus the mean of the truncated sample, we have

$$L = -n \ln [P(c) - P(d+1)] - nm + n\bar{x} \ln m - \ln \left[\prod_1^n x_i! \right]. \tag{7}$$

Differentiating (7) and equating the result to zero, we obtain the estimating equation

$$\frac{1}{n}\frac{dL}{dm} = \frac{\bar{x}}{m} - 1 - \frac{f(c-1) - f(d)}{P(c) - P(d+1)} = 0. \tag{8}$$

For a clearer understanding of how (8) was obtained from (7), the following details are included on determining $dP(c)/dm$. From (2), we have

$$\frac{dP(c)}{dm} = \frac{d\left[\sum_c^\infty \frac{e^{-m}m^x}{x!}\right]}{dm} = \sum_c^\infty \frac{d\left[\frac{e^{-m}m^x}{x!}\right]}{dm},$$

$$= \sum_c^\infty \left[\frac{e^{-m}xm^{x-1} - e^{-m}m^x}{x!}\right],$$

$$= \sum_c^\infty \frac{e^{-m}m^{x-1}}{(x-1)!} - \sum_c^\infty \frac{e^{-m}m^x}{x!},$$

$$= \sum_{c-1}^\infty \frac{e^{-m}m^x}{x!} - \sum_c^\infty \frac{e^{-m}m^x}{x!} = P(c-1) - P(c),$$

and finally

$$\frac{dP(c)}{dm} = f(c-1). \tag{9}$$

With the aid of a set of Poisson tables such as those of Molina [3], equation (8) can be solved for the required estimate, $\hat{m}$, by elementary iterative procedures, one of which is illustrated in Section 7.

Singly truncated on the left. For a sample that is singly truncated on the left, the estimating equation (8) becomes

$$\frac{1}{n}\frac{dL}{dm} = \frac{\bar{x}}{m} - 1 - \frac{f(c-1)}{P(c)} = 0, \tag{10}$$

since here, $d \to \infty$, $\lim_{d\to\infty} f(d) = 0$, and $\lim_{d\to\infty} P(d+1) = 0$. With $c=1$,

estimating equation (10) is applicable in the special case of an unknown number of zero observations, the case considered in [2].

Singly truncated on the right. In this case, $c=0$, $f(c-1)=0$, and $P(c)=1$. Accordingly the estimating equation becomes

$$\frac{1}{n}\frac{dL}{dm}=\frac{\bar{x}}{m}-1+\frac{f(d)}{1-P(d+1)}=0. \tag{11}$$

4. CENSORED SAMPLES—NUMBER UNMEASURED OBSERVATIONS IN EACH TAIL KNOWN

Doubly censored. Let n_1 and n_2 be the number of unmeasured observations in the left and right tails respectively and let n be the number of measured observations for which $c \leqq x \leqq d$. The likelihood function for a sample of this type drawn from the population (1) is

$$\begin{aligned} &P(x_1, x_2, \cdots, x_{n+n_1+n_2}) \\ &\qquad = K[1-P(c)]^{n_1}\left[\frac{e^{-nm}m^{\sum_1^n x_i}}{x_1!\cdots x_n!}\right][P(d+1)]^{n_2}, \end{aligned} \tag{12}$$

where K is a constant, and other symbols are as previously defined. Taking logarithms of (12), differentiating with the aid of (9) and equating to zero, we obtain the estimating equation

$$\frac{1}{n}\frac{dL}{dm}=\frac{\bar{x}}{m}-1-\frac{n_1}{n}\left[\frac{f(c-1)}{1-P(c)}\right]+\frac{n_2}{n}\left[\frac{f(d)}{P(d+1)}\right]=0. \tag{13}$$

Singly censored on the left. In this case $n_2=0$, and the estimating equation (13) becomes

$$\frac{1}{n}\frac{dL}{dm}=\frac{\bar{x}}{m}-1-\frac{n_1}{n}\left[\frac{f(c-1)}{1-P(c)}\right]=0. \tag{14}$$

When $c=1$, a singly censored sample with the number of unmeasured observations known is actually a complete rather than a restricted sample since n_1 is simply the number of zeros in the sample and the total sample size is $n+n_1$. In this case, equation (14) becomes

$$\frac{\sum x_i}{nm}-1-\frac{n_1}{n}\left[\frac{f(0)}{f(0)}\right]=0,$$

and

$$m=\frac{\sum x_i}{n+n_1}=\bar{x},$$

which agrees with equation (3) for a complete sample.

Singly censored on the right. In this instance, $n_1 = 0$, and the estimating equation (13) becomes

$$\frac{1}{n}\frac{dL}{dm} = \frac{\bar{x}}{m} - 1 + \frac{n_2}{n}\left[\frac{f(d)}{P(d+1)}\right] = 0. \tag{15}$$

This is recognized as the appropriate estimating equation for the case in which all sample observations for which $x > d$ have been pooled. Tippett's estimator (loc. cit.) applied to this case.

5. CENSORED SAMPLES—TOTAL NUMBER UNMEASURED OBSERVATIONS KNOWN BUT NOT THE NUMBER IN EACH TAIL SEPARATELY

Let c and d designate the terminals as in Sections 3 and 4. Let n be the number of measured observations for which $c \leq x \leq d$, and n_0 the combined number of unmeasured observations in the two tails. The likelihood function for a sample of this type from the population (1) is then

$$\begin{aligned} &P(x_1, x_2, \cdots, x_{n+n_0}) \\ &\qquad = K[1 - P(c) + P(d+1)]^{n_0}\left[\frac{e^{-nm}m^{\sum_1^n x_i}}{x_1! \cdots x_n!}\right]. \end{aligned} \tag{16}$$

Taking logarithms, differentiating and equating to zero, we have the estimating equation

$$\frac{1}{n}\frac{dL}{dm} = \frac{\bar{x}}{m} - 1 + \frac{n_0}{n}\left[\frac{f(d) - f(c-1)}{1 - P(c) + P(d+1)}\right] = 0. \tag{17}$$

We note that the singly censored cases in this instance are identical with those of Section 4. When censored on the left, $n_0 = n_1$, $\lim_{d\to\infty} f(d) = 0$, and $\lim_{d\to\infty} P(d+1) = 0$. Accordingly, (17) assumes the same form as (14). When censored on the right only, $n_0 = n_2$, $c = 0$, $f(c-1) = 0$, $P(c) = 1$, and (17) reduces to the same form as (15).

6. VARIANCE OF ESTIMATES

Since maximum likelihood estimation has been employed, the asymptotic variance of $\hat{m}$ can be expressed as

$$\operatorname{Var}(m) = -\left[\frac{d^2L}{dm^2}\right]^{-1}_{m=\hat{m}}. \tag{18}$$

Second derivatives for the various cases considered are given below.

TRUNCATED SAMPLES—NUMBER MISSING OBSERVATIONS UNKNOWN

Doubly Truncated

$$\frac{1}{n}\frac{d^2L}{dm^2} = -\frac{\bar{x}}{m^2} - \left[\frac{f(c-2)-f(c-1)-f(d-1)+f(d)}{P(c)-P(d+1)}\right] + \left[\frac{f(c-1)-f(d)}{P(c)-P(d+1)}\right]^2. \tag{19}$$

Singly Truncated on Left

$$\frac{1}{n}\frac{d^2L}{dm^2} = -\frac{\bar{x}}{m^2} - \left[\frac{f(c-2)-f(c-1)}{P(c)}\right] + \left[\frac{f(c-1)}{P(c)}\right]^2. \tag{20}$$

Singly Truncated on Right

$$\frac{1}{n}\frac{d^2L}{dm^2} = -\frac{\bar{x}}{m^2} + \left[\frac{f(d-1)-f(d)}{1-P(d+1)}\right] + \left[\frac{f(d)}{1-P(d+1)}\right]^2. \tag{21}$$

CENSORED SAMPLES—NUMBER OBSERVATIONS IN EACH TAIL KNOWN

Doubly Censored

$$\frac{1}{n}\frac{d^2L}{dm^2} = -\frac{\bar{x}}{m^2} - \frac{n_1}{n}\left[\frac{f(c-2)-f(c-1)}{1-P(c)} + \left(\frac{f(c-1)}{1-P(c)}\right)^2\right] + \frac{n_2}{n}\left[\frac{f(d-1)-f(d)}{P(d+1)} - \left(\frac{f(d)}{P(d+1)}\right)^2\right]. \tag{22}$$

Singly Censored on Left

$$\frac{1}{n}\frac{d^2L}{dm^2} = -\frac{\bar{x}}{m^2} - \frac{n_1}{n}\left[\frac{f(c-2)-f(c-1)}{1-P(c)} + \left(\frac{f(c-1)}{1-P(c)}\right)^2\right]. \tag{23}$$

Singly Censored on Right

$$\frac{1}{n}\frac{d^2L}{dm^2} = -\frac{\bar{x}}{m^2} + \frac{n_2}{n}\left[\frac{f(d-1)-f(d)}{P(d+1)} - \left(\frac{f(d)}{P(d+1)}\right)^2\right]. \tag{24}$$

CENSORED SAMPLES—COMBINED NUMBER OBSERVATIONS IN TAILS KNOWN

$$\frac{1}{n}\frac{d^2L}{dm^2} = -\frac{\bar{x}}{m^2} + \frac{n_0}{n}\left[\frac{f(d-1) - f(d) - f(c-2) + f(c-1)}{1 - P(c) + P(d+1)} - \left(\frac{f(d) - f(c-1)}{1 - P(c) + P(d+1)}\right)^2\right]. \tag{25}$$

Derivatives given in this section can be evaluated with the aid of Molina's Tables as previously mentioned.

7. PRACTICAL APPLICATIONS

For each case considered in this paper, the estimating equation can be solved for $\widehat{m}$ without too much difficulty by a simple inverse interpolative process, provided a set of Poisson tables such as those of Molina [3] are available. As a first approximation to $\widehat{m}$ the sample mean, $\bar{x}$, or some obvious modification thereof, will prove satisfactory in many instances. Where applicable, estimates given by Moore (loc. cit.) or by Rider (loc. cit.) may provide closer first approximations. The illustrations which follow serve to clarify these points.

The data of Table 1, due to Rutherford and Geiger [6], will be used to illustrate how estimates are calculated from samples. These data concern the number of α particles observed in an eighth of a minute time interval. Observations were recorded for 2608 such intervals, with x designating the number of particles observed during an interval, and $f_0(x)$ designating the frequency or number of intervals during which these observations were made. Observations of nine or more particles per time interval were pooled. Various cases considered in this paper are illustrated with these same data by appropriately changing basic assumptions regarding the sample.

TABLE 1

Particles per Interval, x	0	1	2	3	4	5	6	7	8	9 and over	Total
Number of Intervals, $f_0(x)$	57	203	383	525	532	408	273	139	45	43	2608

Illustration 1

Sample singly censored on the right. We first use the maximum information provided by the sample which includes $n = 2565$ measured

observations and $n_2=43$ unmeasured or pooled observations in the sample right tail. The terminus is at $d=8$, and $\sum_{x=0}^{8} xf_0(x)=9683$. Accordingly, $\bar{x}=9683/2565=3.7750$. The appropriate estimating equation is (15) which we solve, using as a first approximation, $m_1=\bar{x}=3.8$ (rounded off). From Molina's tables, we have $f(8, 3.8)=0.024123$ and $P(9, 3.8)=0.015984$. On substituting these values in (15) we obtain

$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.8} = \frac{3.7750}{3.8} - 1 + \frac{43}{2565}\left[\frac{0.024123}{0.015984}\right] = +0.01872.$$

Similarly, we compute

$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.9} = \frac{3.7750}{3.9} - 1 + \frac{43}{2565}\left[\frac{0.026869}{0.018533}\right] = -0.00775.$$

To determine the required value, $\hat{m}$, for which $(1/n)dL/dm=0$, we interpolate linearly[1] as summarized below.

m	$\frac{1}{n}\frac{dL}{dm}$
3.800	+0.01872
3.871	0.00000
3.900	−0.00775

Thus our estimate is $\hat{m}=3.871$, and using (18) and (24), we compute

$$\sigma_{\hat{m}} = \sqrt{V(\hat{m})} \sim 0.04.$$

Illustration 2

Sample singly truncated on the right. Here, we neglect n_2 and assume that the number of observations in the truncated tail is unknown. Otherwise the sample remains the same as for illustration 1. Estimating equation (11) is applicable in this instance, and on substituting the necessary values, we have

$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.8} = \frac{3.7750}{3.8} - 1 + \frac{0.024123}{1-0.015984} = +0.01794,$$

$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.9} = \frac{3.7750}{3.9} - 1 + \frac{0.026869}{1-0.018533} = -0.00468.$$

[1] More precise interpolation formulas involving second and higher order differences might be required to give the desired accuracy in some applications.

Interpolating as in the previous illustration, we have $\hat{m}=3.879$, and from (18) and (21), $\sigma_{\hat{m}}\sim 0.04$.

Illustration 3

Doubly truncated sample. For this illustration, we arbitrarily eliminate the first two classes of Table 1 in addition to the pooled classes for measurements greater than 8, and assume the number of missing observations to be unknown. In this instance the terminals are $c=2$ and $d=8$. Completing the sample summary, we have $\sum_{x=2}^{8} xf_0(x)=9480$, $n=2305$, and $\bar{x}=9480/2305=4.112798$. The appropriate estimating equation is (8) and from Molina's tables, we find $f(1, 3.8)=0.085009$, $f(8, 3.8)=0.024123$, $P(2, 3.8)=0.892620$, $P(9, 3.8)=0.015984$, $f(1, 3.9)=0.078943$, $f(8, 3.9)=0.026869$, $P(2, 3.9)=0.900815$, and $P(9, 3.9)=0.018533$. On substituting these values in (8) we have

$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.8} = \frac{4.112798}{3.8} - 1 - \left[\frac{0.085009 - 0.024123}{0.892620 - 0.015984}\right]$$
$$= +0.01286,$$
$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.9} = \frac{4.112798}{3.9} - 1 - \left[\frac{0.078943 - 0.026869}{0.900815 - 0.018533}\right]$$
$$= -0.00446.$$

Interpolating, we have $\hat{m}=3.874$, and from (18) and (19), $\sigma_{\hat{m}}\sim 0.05$.

Illustration 4

Sample doubly censored with number unmeasured observations in each tail known. Data for this illustration are the same as for illustration 3 with the added information that $n_1=260$, and $n_2=43$. The applicable estimating equation is (13), and on making necessary substitutions, we obtain

$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.8} = \frac{4.112798}{3.8} - 1 - \frac{260}{2305}\left[\frac{0.085009}{1 - 0.892620}\right]$$
$$+ \frac{43}{2305}\left[\frac{0.024123}{0.015984}\right] = +0.02117,$$
$$\frac{1}{n}\frac{dL}{dm}\bigg|_{m=3.9} = \frac{4.112798}{3.9} - 1 - \frac{260}{2305}\left[\frac{0.078943}{1 - 0.900815}\right]$$
$$+ \frac{43}{2305}\left[\frac{0.026869}{0.018533}\right] = -0.00817.$$

Interpolating, we find $\hat{m}=3.872$, and from (18) and (22), $\sigma_{\hat{m}}\sim 0.04$.

Illustration 5

Sample doubly censored with total number unmeasured observations known, but not number in each tail separately. For this illustration we assume the knowledge that $n_0=303$, but that n_1 and n_2 separately are unknown. Otherwise, the data remain the same as for illustrations 3 and 4. Estimating equation (17) is applicable, and after making the necessary substitutions, we have

$$\left.\frac{1}{n}\frac{dL}{dm}\right|_{m=3.8} = \frac{4.112798}{3.8} - 1 + \frac{303}{2305}\left[\frac{0.024123 - 0.085009}{1 - 0.892620 + 0.015984}\right]$$

$$= +0.01744,$$

$$\left.\frac{1}{n}\frac{dL}{dm}\right|_{m=3.9} = \frac{4.112798}{3.9} - 1 + \frac{303}{2305}\left[\frac{0.026869 - 0.078943}{1 - 0.900815 + 0.018533}\right]$$

$$= -0.00359.$$

Interpolation yields $\hat{m}=3.883$, and from (18) and (25), $\sigma_{\hat{m}}\sim 0.05$.

We dispense with illustrations of the remaining sample types since solution of applicable estimating equations can be accomplished in the same manner as in the five illustrations presented. To solve estimating equations for any of the cases considered in this paper, standard iterative procedures such as Newton's method could be employed, but for simplicity and ease of application, the interpolative procedure illustrated seems preferable.

REFERENCES

[1] Bliss, C. I., "Estimation of the mean and its error from incomplete Poisson distributions," Connecticut Agricultural Experiment Station Bulletin 513 (1948), 12 p.

[2] David, F. N., and Johnson, N. L., "The truncated Poisson," *Biometrics*, 8 (1952), 275–85.

[3] Molina, E. C., *Poisson's Exponential Binomial Limit*, D. Van Nostrand Co., Inc., (1942).

[4] Moore, P. G., "The estimation of the Poisson parameter from a truncated distribution," *Biometrika*, 39 (1952), 247–51.

[5] Rider, Paul R., "Truncated Poisson distributions," *Journal of the American Statistical Association*, 48 (1953), 826–30.

[6] Rutherford, E., and Geiger, Hans, "The probability variations in the distribution of α particles," *Phil. Magazine Series 6*, 20 (1910), 698.

[7] Tippett, L. H. C., "A modified method of counting particles," *Proceedings Royal Society, Series A*, 137 (1932), 434–46.

TABLES OF THE EXPECTED VALUE OF 1/X FOR POSITIVE BERNOULLI AND POISSON VARIABLES

Edwin L. Grab and I. Richard Savage
National Bureau of Standards

SUMMARY OF TABLES

The random variable X is said to have a positive Bernoulli distribution [11][1] if the probability that $X=x$ is equal to $\binom{n}{x}p^x q^{n-x}(1-q^n)^{-1}$ for $x=1, 2, \cdots, n$ where $q=1-p$ and $0<p\leqq 1$. Similarly the variable X is said to have a positive Poisson distribution if the probability that $X=x$ is equal to $e^{-m}(1-e^{-m})^{-1}m^x/x!$ for $x=1, 2, \cdots$, and $m>0$.

Table I gives the values of $E(1/X|n, p)$ to five decimal places where

$$(1) \qquad E(1/X \mid n, p) = (1-q^n)^{-1}\sum_{x=1}^{n}\binom{n}{x}x^{-1}p^x q^{n-x} \qquad (q = 1-p)$$

for the following values of the parameters:

$$n = 2(1)20; \quad p = .01, .05(.05).95, .99$$
$$n = 21(1)30; \quad p = .01, .05(.05).50.$$

Table II gives values of $E(1/X|n)$ to five decimal places, where

$$(2) \qquad E(1/X \mid m) = e^{-m}(1-e^{-m})^{-1}\sum_{x=1}^{\infty} m^x/(x!x),$$

for the following values of the parameter:

$$m = .01, .05(.05)1.0(.1)2.0(.2)5.0(.5)7.0(.1)10(2)20.$$

PREPARATION OF TABLES

Table I was originally prepared directly from (1) using *Tables of the Binomial Probability Distribution* [9]. Subsequently the table was checked by using the recurrence relation

$$(3) \qquad E(1/X \mid n+1, p) = \frac{1}{n+1} + q\,\frac{1-q^n}{1-q^{n+1}}\,E(1/X \mid n, p).$$

Table II was prepared directly from (2) making use of tables of the Poisson distribution [4], [5], [6]. Some of the values were checked by

[1] This article by Stephan contains much material of interest related to the contents of this paper In particular it gives a precise formulation of the mathematical situation in which these tables are applicable in sampling problems. Also it presents many more results of interest.

the use of

$$E(1/X \mid m) = [Ei(m) - \gamma - \log_e m]e^{-m}/(1 - e^{-m}), \tag{4}$$

making use of [7], [8], and [10].

The entries in Tables I and II were obtained by two independent computations to insure five-decimal accuracy.

USE OF TABLES

The following situation often arises in sampling problems.

An observation y_i $(i=1, \cdots, x)$ is made on each of x individuals and the average,

$$Y = (y_1 + y_2 + \cdots + y_x)/x,$$

is computed.

If the y_i's are independent observations on a random variable with mean value μ and variance σ^2, we find that the mean value of Y is μ. However, if x, the sample size, is a random variable, then the variance of Y is not σ^2/x but is $\sigma^2 E(1/X)$. Thus one use of the tables is in finding the variances of means when the sample sizes are random variables with either positive Bernoulli or Poisson distributions. Extensive discussions of the above situation can be found in [2], [3], [11], and [12].

Some typical situations where X has a positive Bernoulli or Poisson distribution are:

1) In estimating the average number of acres per farm planted with cotton, of those farms of a sample having any cotton planted.
2) In estimating the average weight of animals that will survive a certain experiment, where the probability of an animal dying is constant for each animal and independent of the other animals.
3) In estimating the average cost of fires in a certain city by examining the cost of all fires that occurred in a short time interval.

In using the tables one must have good estimates of p or m. However, the above examples are typical in that they cover situations where one is likely to have good estimates of the important quantities, i.e., proportion of farms growing cotton, the lethality of an experiment, and the average number of fire alarms per day.

In some sampling problems the sample size follows other discrete distributions, such as the hypergeometric. Tables of $E(1/X)$ for this distribution would be difficult to prepare since the distribution itself is so poorly tabulated and since there are three parameters involved. Hence, if in dealing with this distribution one feels that the binomial or Poisson approximations are not adequate, then one could perform the

desired computations for the specific situations at hand. For some discrete distributions the formula for $E(1/X)$ is simple, as for example the Geometric distribution.

APPROXIMATIONS

Stephan [11] gives several methods for computing (1), but none of these gives a simple approximation for the entire range of parameters covered by Table I. His approximations are advantageous to use for larger values of n than those covered by the present tables (See his examples).

Finkner [3] suggests for large values of np that the following relationship will hold:

$$1/np < E(1/X \mid n, p) < 1/(np - 1). \tag{5}$$

In preparing Table I it was noted for large values of np that a very good approximation to $E(1/X|n, p)$ is given by

$$1/(np - q). \tag{6}$$

We are told by one of the referees that (6) also appears in an unpublished manuscript of W. A. Hendricks. The bounds in (5) and (6) suffer from the disadvantage that there is no theory as to when they are good approximations. On the other hand it is clear that sometimes they are poor since $1/(np-1)$ can take on negative values and $1/np$, $1/(np-q)$, and $1/(np-1)$ can take on values larger than one.

As an immediate consequence of the Schwarz inequality we find for any positive random variable that the following inequality is valid:

$$E(1/X) \geqq 1/E(X). \tag{7}$$

This inequality in the case of the positive Bernoulli and Poisson distributions becomes

$$E(1/X \mid n, p) \geqq (1 - q^n)/np \tag{8}$$

and

$$E(1/X \mid m) > (1 - e^{-m})/m, \tag{9}$$

respectively. In (8) the equality holds if either $n=1$ or $p=1$.

Using the inequality

$$1/x \leqq 1/(x + 1) + 3/(x + 1)(x + 2), \tag{10}$$

which is valid for $x \geqq 1$ we have for random variables which cannot take on values less than one the following inequality:

$$E(1/X) \leqq E[1/(X+1)] + 3E[1/(X+1)(X+2)]. \tag{11}$$

In the case where X has a positive Bernoulli distribution this gives

$$\begin{aligned} E(1/X \mid np) \leqq{} & [p(n+1)(1-q^n)]^{-1}[(1-P(1 \mid n+1, p)) \\ & + 3(1-P(2 \mid n+2, p))/(n+2)p] \end{aligned} \tag{12}$$

where $P(i|r, p)$ is the probability that a Bernoulli variable with parameters r and p will be less than or equal to i. Finally when X has a positive Poisson distribution we obtain

$$\begin{aligned} E(1/X \mid m) \leqq{} & [(1-e^{-m})m]^{-1}[(1-P(1 \mid m)) \\ & + 3(1-P(2 \mid m))/m], \end{aligned} \tag{13}$$

where $P(i|m)$ is the probability that a Poisson variable with parameter m will be less than or equal to i.

This paper is primarily concerned with obtaining exact values of $E(1/X|n, p)$ and $E(1/X|m)$. Techniques for finding the limiting distributions and moments of reciprocals of positive Bernoulli and Poisson random variables are available in sections 27.7 and 28.4 or [1].

INTERPOLATION AND EXTRAPOLATION

One entry in each column of Table I bears an asterisk. For values of p equal to or larger than the one corresponding to the entry with an asterisk, the approximation $1/(np-q)$ is accurate to at least two decimal places and usually to two significant figures; and, in general, if $np>10$, it has been found that $1/(np-q)$ gives at least two-place accuracy. These statements are empirical. Although it has not been possible to derive these facts mathematically they have been observed in many computations.

If $n \leqq 10$ and p is above the asterisk, linear interpolation gives results accurate to two decimal places.

For the cases not covered in the two preceding paragraphs more complicated interpolation formulations might be advantageous to use. In particular when np or m are greater than 5 it has been noted that interpolation of the functions $pE(1/X|n, p)$ and $mE(1/X|m)$ give better results than linear interpolation.

For values of n greater than 30, it has been found that if $np<2$ one may set $np=m$ and use Table II. This method seems always to yield at least two significant figures.

Formula (3) is easy to apply and may be found useful for extending Table I to other values of n as needed.

ACKNOWLEDGMENT

Thanks are due Lola S. Deming for her careful checking of the tables, and for correction of a number of errors found in the original values.

REFERENCES

[1] Cramer, H., *Mathematical Methods of Statistics.* Princeton, N. J.: Princeton University Press, 1951.
[2] Deming, W. E., *Some Theory of Sampling.* New York: John Wiley and Sons, Inc., 1950, 449–54.
[3] Finkner, A. L., "Further investigation on the theory and application of sampling for scarcity items," *Institute of Statistics, University of North Carolina, Mimeo. Series* 30.
[4] Fry, T. C., *Probability and It's Engineering Uses*, D. Van Nostrand Company, Inc. 458–62 (1928).
[5] Kitagawa, T., *Tables of Poisson Distribution*, Biafukan, Tokyo, Japan, (1952).
[6] Molina, E. C., *Poisson's Exponential Binomial Limit*, D. Van Nostrand Company, Inc., (1947).
[7] National Bureau of Standards, *Tables of Sine, Cosine, and Exponential Integrals*, 2, MT6, U. S. Government Printing Office, Washington, D. C.
[8] National Bureau of Standards, *Tables of Natural Logarithms*, MT7, MT9, MT10, MT12, U. S. Government Printing Office, Washington, D. C.
[9] National Bureau of Standards, *Table of the Binomial Probability, Distribution*, Applied Mathematics Series 6.
[10] National Bureau of Standards, *Tables of the Exponential Function* e^x, Applied Mathematics Series 14.
[11] Stephan, F. F., "The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate," *Annals of Mathematical Statistics*, 16 (1945), 50–61.
[12] Horvitz, D. G., *Ratio method of estimation in sample surveys*, Unpublished thesis, Iowa State College, 1953.

TABLE 1

$E(1/X \mid n, p) = (1-q^n)^{-1} \sum_{x=1}^{n} \binom{n}{x} x^{-1} p^x q^{n-x} \quad (q = 1-p)$

p \ n	2	3	4	5	6	7	8	9	10	11
.01	.99749	.99498	.99247	.98997	.98747	.98497	.98247	.97998	.97749	.97500
.05	.98718	.97444	.96178	.94920	.93671	.92431	.91200	.89979	.88767	.87565
.10	.97368	.94772	.92214	.89696	.87220	.84787	.82400	.80060	.77768	.75526
.15	.95946	.91983	.88117	.84357	.80708	.77176	.73763	.70475	.67312	.64277
.20	.94444	.89071	.83898	.78940	.74210	.69715	.65461	.61450	.57682	.54152
.25	.92857	.86036	.79571	.73489	.67806	.62529	.57657	.53184	.49095	.45371
.30	.91176	.82877	.75158	.68055	.61583	.55736	.50492	.45819	.41674	.38010
.35	.89394	.79594	.70683	.62697	.55629	.49441	.44067	.39429	.35440	.32018
.40	.87500	.76190	.66176	.57474	.50026	.43725	.38436	.34016	.30327	.27243
.45	.85484	.72672	.61676	.52451	.44843	.38637	.33604	.29523	.26204	.23487
.50	.83333	.69048	.57222	.47688	.40132	.34194	.29530	.25847	.22911	.20541
.55	.81034	.65330	.52862	.43241	.35926	.30378	.26142	.22864	.20285	.18217
.60	.78571	.61538	.48645	.39156	.32233	.27147	.23348	.20447	.18177	.16361
.65	.75926	.57697	.44622	.35465	.29037	.24436	.21049	.18477	.16467*	.14854*
.70	.73077	.53837	.40843	.32183	.26305	.22173	.19151	.16956	.15057	.13608
.75	.70000	.50000	.37355	.29311	.23989	.20282	.17570	.15504	.13876	.12560
.80	.66666	.46237	.34188	.26829	.22031*	.18692*	.16238*	.14359*	.12872	.11665
.85	.63043	.42608	.31373	.24704	.20372	.17341	.15101	.13376	.12006	.10892
.90	.59091	.39189	.28915*	.22891*	.18956	.16181	.14118	.12523	.11252	.10216
.95	.54762	.36065*	.26803	.21340	.17734	.15172	.13259	.11774	.10589	.09620
.99	.50990*	.33843	.25338	.20253	.16869	.14454	.12645	.11238	.10112	.09192

* For explanation, see section entitled Interpolation and Extrapolation.

TABLE I—(*Continued*)

p \ n	12	13	14	15	16	17	18	19	20
.01	.97252	.97004	.96756	.96509	.96262	.96015	.95769	.95523	.95277
.05	.86373	.85191	.84020	.82859	.81709	.80570	.79443	.78326	.77222
.10	.73334	.71193	.69105	.67069	.65087	.63158	.61282	.59460	.57691
.15	.61370	.58591	.55938	.53411	.51007	.48723	.46556	.44502	.42557
.20	.50856	.47785	.44932	.42284	.39832	.37565	.35470	.33535	.31750
.25	.41990	.38930	.36163	.33667	.31415	.29384	.27552	.25898	.24403
.30	.34780	.31937	.29433	.27228	.25282	.23562	.22038	.20682	.19472
.35	.29081	.26557	.24382	.22502	.20868	.19442*	.18190*	.17085*	.16104*
.40	.24655	.22473	.20619	.19034	.17668	.16482	.15444	.14529	.13717
.45	.21243	.19372	.17795	.16453*	.15299*	.14296	.13419	.12643	.11954
.50	.18601	.16992	.15638	.14486	.13493	.12629	.11870	.11198	.10599
.55	.16530*	.15131*	.13952*	.12945	.12075	.11316*	.10648	.10055	.09525
.60	.14878	.13643	.12600	.11707	.10933	.10255	.09658	.09126	.08650
.65	.13532	.12429	.11493	.10689	.09991	.09379	.08838	.08357	.07925
.70	.12416	.11417	.10568	.09837	.09201	.08643	.08148	.07708	.07312
.75	.11473	.10561	.09783	.09112	.08528	.08014	.07559	.07153	.06788
.80	.10666	.09826	.09108	.08488	.07948	.07472	.07050	.06673	.06335
.85	.09967	.09187	.08521	.07945	.07442	.06999	.06605	.06254	.05938
.90	.09355	.08628	.09006	.07467	.06997	.06582	.06214	.05885	.05588
.95	.08814	.08133	.07550	.07044	.06602	.06212	.05866	.05556	.05278
.99	.08425	.07777	.07221	.06739	.06317	.05946	.05615	.05319	.05053

TABLE I—(Continued)

p \ n	21	22	23	24	25	26	27	28	29	30
.01	.95031	.94786	.94541	.94297	.94053	.93809	.93565	.93322	.93079	.92837
.05	.76128	.75047	.73977	.72920	.71874	.70841	.69820	.68811	.67814	.66830
.10	.55975	.54312	.52700	.51140	.49630	.48171	.46760	.45397	.44081	.42811
.15	.40718	.38980	.37338	.35788	.34326	.32947	.31647	.30420	.29264	.28174
.20	.30203	.28583	.27180	.25885	.24688	.23582	.22558	.21609	.20729	.19911
.25	.23049	.21822	.20707	.19692	.18765	.17917	.17140	.16425	.15766	.15157
.30	.18389	.17416	.16537	.15742	.15019	.14359	.13755	.13199	.12688	.12215
.35	.15229*	.14444*	.13736*	.13095*	.12512*	.11979*	.11490*	.11040*	.10624*	.10239*
.40	.12992	.12341	.11752	.11218	.10731	.10285	.09874	.09496	.09146	.08821
.45	.11336	.10780	.10277	.09819	.09400	.09016	.08663	.08336	.08033	.07752
.50	.10061	.09576	.09136	.08735	.08367	.08030	.07719	.07431	.07164	.06915

TABLE II

$E(1/X|\mathrm{m}) = e^{-m}(1-e^{-m})^{-1}\sum_{x=1}^{\infty} m^x/(x!x)$

m	$E(1/X\|m)$	m	$E(1/X\|m)$
.01	.99750	1.9	.59351
.05	.98754	2.0	.57659
.10	.97514	2.2	.54417
.15	.96282	2.4	.51361
.20	.95058	2.6	.48488
.25	.93842	2.8	.45792
.30	.92636	3.0	.43268
.35	.91435	3.2	.40909
.40	.90244	3.4	.38707
.45	.89062	3.6	.36654
.50	.87889	3.8	.34742
.55	.86725	4.0	.32963
.60	.85571	4.2	.31308
.65	.84426	4.4	.29770
.70	.83292	4.6	.28340
.75	.82166	4.8	.27012
.80	.81052	5.0	.25777
.85	.79948	5.5	.23055
.90	.78854	6.0	.20779
.95	.77771	6.5	.18866
1.00	.76699	7.0	.17249
1.10	.74587	8	.14689
1.20	.72520	9	.12776
1.30	.70499	10	.11302
1.40	.68523	12	.09190
1.50	.66594	14	.07749
1.60	.64712	16	.06702
1.70	.62878	18	.05906
1.80	.61019	20	.05280

STATISTICAL ABSTRACTS

This section, an experiment, will present abstracts of articles on statistical methods. Such articles appear in numerous journals and it is the aim of this section to call attention to them by brief summaries. The object of each abstract is not a critical review of an article but a statement in as nonmathematical terms as possible of the statistical problem considered, the character of the methods used to solve it, and the results obtained.

Certain journals which normally contain articles of statistical interest will be abstracted regularly. These are *American Journal of Public Health, Annals of Eugenics, Annals of Mathematical Statistics, Biometrics, Biometrika, Calcutta Statistical Bulletin, Econometrica, Human Biology, Journal of Agricultural Sciences, Journal of the Royal Statistical Society—Series B, Psychometrika, Sankya,* and *Sociometry.* Most articles on statistical methods appearing in these journals in 1953 or later will be abstracted. Papers on statistical methods published elsewhere will be included as they come to the attention of the Abstracts Editor; readers are invited to submit to him (at the address below) abstracts or suggestions of papers for abstracting.

The usefulness of the section will depend on the thoroughness of coverage, the quality, and the style of the abstracts. Criticisms and suggestions are invited. The Department of Statistics of the University of North Carolina has accepted the responsibility for the section and will depend not only on the faculty and graduate students of the Institute of Statistics but also on correspondents from other universities, government, and industry.

All communications concerning this section should be addressed to the Abstracts Editor, Professor George E. Nicholson, Jr., Chairman of the Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Abelson, R. P., "A note on the Neyman-Johnson technique," *Psychometrika,* 18 (1953), 213–18.

In a multivariate prediction problem, for what values of the predictor variables will two groups differ significantly on the criterion variable? The region of significance of the Neyman-Johnson technique is defined as the set of points of the predictor space where one group is significantly better than the other on the criterion variable. These latter authors have provided an analytic definition of this region for the case of three predictor variables. The present author generalizes the solution to any number of predictors. A ratio approximating the generalized region of significance is proposed and this ratio is shown to be asymptotically equivalent to the expression obtained by Neyman and Johnson. The derivation given by the author is a straightforward application of the critical ratio principle to the difference between predicted criterion scores for the two groups. The order of the approximation of the ratio given here to the Neyman-Johnson ratio is that of the order with which the Beta distribution is approximated by the F distribution. B. J. WINER, *University of North Carolina.*

Bartlett, M. S., "The statistical significance of odd bits of information," *Biometrika,* 39 (1952), 228–37.

The author presents a method of pooling information from n independent events to test a given hypothesis, H. If each event is a dichotomy, the total information if $i_n = -\Sigma \log p_i$, where p_i is the probability of the event occurring, given H. The expected value and variance of i_n are $E(i_n) = I_n = -\Sigma_i(p_i \log p_i + q_i \log q_i)$; $\sigma_{i_n}^2 = \Sigma_i p_i q_i [\log (p_i/q_i)]$. All logarithms are to the base e.

H is tested by considering the ratio $(\hat{\imath}_n - I_n)/\sigma_{\hat{\imath}_n}$ as a normal deviate. The results are extended to a multichotomy. An example is presented. Some observations are made on the use of this method for certain types of dependent observations. A brief discussion is given on the construction of confidence limits when the estimator is the ratio of random variates. R. L. ANDERSON, *North Carolina State College.*

Bliss, C. I., "Fitting the negative binomial distribution to biological data"; Fisher, R. A., "Note on the efficient fitting of the negative binomial," *Biometrics*, 9 (1953), 176–200.

See Fisher, R. A.

Cox, D. R., "Estimation by double sampling," *Biometrika*, 39 (1952), 217–27.

Double sampling methods are developed to obtain a large sample estimator, t, of some parameter, θ, both for known and unknown population variance. The variance of t is assumed to be some function of θ, $a(\theta)$, given in advance. The results are applied to the estimation of normal and binomial means, when $a(\theta)=a$ or $a\theta^n$. Estimation by confidence intervals and a combined estimation and testing procedure are also considered. R. L. ANDERSON, *North Carolina State College.*

Dixon, W. J., "Processing data for outliers," *Biometrics*, 9 (1953), 74–89.

The problem considered is how the mean, μ, and standard deviation, σ, should be estimated where N observations from a normal population $N(\mu, \sigma)$ may actually contain some small unknown proportion γ of observations from a "contaminating" population with a different standard deviation, $N(\mu, \lambda^2\sigma^2)$, or from one with a different mean, $N(\mu+\lambda\sigma, \sigma^2)$. The results in the paper are stated for sample sizes $N=5$ and $N=15$. The numerical results are based in large part on experimental sampling. The behavior of mean and median (as estimates of μ) and of S^2 (estimate of σ^2), and of s and the range (as estimators of σ) is investigated for various values of λ and γ. The effect of the estimators of processing the data for outlying observations using various levels of significance is explored. For the whole range of λ and γ, and for sample sizes $N=5$ and $N=15$, recommendations are made as to the level of significance to use in processing data for outliers, and which estimates to employ. LINCOLN MOSES, *Stanford University.*

Draper, J., "Properties of distributions resulting from certain simple transformations of the normal distribution," *Biometrika*, 39 (1952), 290–301.

Given a non-normal variate, x, to be transformed to a normal z with unit variance. The following transformations are considered: $S_L: z=\gamma+\delta \log (x-\xi)$; $S_U: z=\gamma+\delta \sinh^{-1} y$; $S_B: z=\gamma+\delta \log [y/(1-y)]$, where $y=(x-\xi)/\lambda$. Methods of estimating the parameters, γ, δ, ξ, and λ are given for S_U. Several examples are given, including its use for normalizing t and non-central t. A quadrature formula is suggested for estimating the moments of the S_B system. The author states that the calculation of the parameters of the S_L system "presents no difficulty using the first three moments of the distribution of x." R. L. ANDERSON, *North Carolina State College.*

Fisher, R. A., "Note on the efficient fitting of the negative binomial"; Bliss, C. I., "Fitting the negative binomial distribution to biological data," *Biometrics*, 9 (1953), 176–200.

The negative binomial distribution is a two parameter distribution for a discrete random variable which gives the probability of x occurrences of an event in a sampling unit as $[(k+x-1)!]/[x_i!(k-1)!][p^x/(1+p)^{k+x}]$. For limiting values of the parameters this distribution yields the Poisson or the Fisher logarithmic series as special cases. The distribution is of wide application in biological sciences, having been used to describe insect populations, distribution of bacterial clumps, accident rates, etc. This paper discusses various models leading to the negative binomial probability distribution, and to various alternative distributions more or less related to it. The mean of the distribution (which is pk) is efficiently estimated by the sample mean $\bar{x}$. The parameter k has been estimated by the method of moments, or from the number of zero occurrences. Fisher here gives the maximum likelihood estimates of the parameters together with a convenient arithmetical scheme of computation. Tests of fit to the model are discussed. Procedures are illustrated with numerical examples. LINCOLN MOSES, *Stanford University.*

Grundy, P. M., "The fitting of grouped truncated and grouped censored normal distributions," *Biometrika*, 39 (1952), 252–59.

A distribution is said to be censored if the frequency of observations in the truncated

region is known although their values are unknown. A process involving adjusted sample moments, which is used in connection with published tables, is shown to be equivalent to maximum likelihood estimation. In the special case when the group intervals are equal, approximate formulas for the adjusted moments become particularly simple. The accuracy of the approximations is indicated. Information and covariance matrices are studied from the standpoint of effects of grouping. A numerical example illustrates the principles. T. W. HORNER, *North Carolina State College*.

Gupta, A. K., "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, 39 (1952), 260–73.

A sample may be censored in two ways: I. observations below or above a given point may be censored; II. the $(n-k)$ smallest or greatest observations out of a sample of size n may be censored. The author was concerned with estimating the mean and standard deviation of a normal population from a type II censored sample. Tables were given which facilitate the computation of the maximum likelihood estimates and their asymptotic variances and covariances. Since the maximum likelihood estimates may be biased for small n, the best linear unbiased estimate was derived. Coefficients for finding the best linear estimate of the mean and the standard deviation from censored data for $n \leq 10$ are given. An alternative unbiased linear estimate, which has great efficiency, was proposed for n slightly larger than 10. Examples illustrate the three methods of estimation. T. W. HORNER, *North Carolina State College*.

Hyrenius, H., "Sampling from bivariate non-normal universes by means of compound normal distributions," *Biometrika*, 39 (1952), 238–46.

The effect of non-normality on estimates of the correlation and regression coefficients and their variances is studied by considering each sample to be from a different bivariate normal universe. Only inequality in the means for the different universes is considered in this paper. R. L. ANDERSON, *North Carolina State College*.

Jacob, Walter C., "Split-plot half-plaid squares for irrigation experiments," *Biometrics*, 9 (1953), 157–75.

The objectives of the experiment were to study the effects of nitrogen, phosphate, and potash (each at 3 levels) on the yield (in pounds) of U. S. No. 1 tubers for three varieties of potatoes in the presence or absence of irrigation; altogether about two and a half acres were available in two blocks of equal size. All combinations of varieties, fertilizers, and irrigation imply 162 different treatments. Since the main effect of irrigation was of little interest (but its interactions were of interest) the two blocks were each divided into two plots, one irrigated, and one not. The 81 treatment combinations to be applied to the four split plots were arranged as a 9×9 quasi-latin square by confounding portions of second and third order interactions with the rows and columns of the field. This removal of row and column sum of squares resulted in about 100% gain in precision. Numerical analysis of the data, and the interpretation are given. LINCOLN MOSES, *Stanford University*.

Johnson, N. L., "Approximations to the probability integral of the distribution of range," *Biometrika*, 39 (1952), 417–19.

Given a random sample of n from $F(x)$. Approximate formulas for the probability of the range not exceeding w are given. Comparisons are made with exact probabilities and significance levels for $F(x)$ normal. R. L. ANDERSON, *North Carolina State College*.

Kaplan, E. L., "Tensor notation and the sampling cumulants of k-statistics," *Biometrika*, 39 (1952), 319–23.

Concise formulas are given for sampling from infinite populations. It is shown that results for multivariate distributions are mild generalizations of those for univariate relations. R. L. ANDERSON, *North Carolina State College*.

Kimball, A. W., "The fitting of multi-hit survival curves," *Biometrics*, 9 (1953), 201–11.

Let a population of organisms be exposed to a dose of radiation, x. Suppose that an organism loses its viability if and only if all of n "sensitive units" in the organism are inactivated, or hit. Further, assume that the probability of any one unit being hit is e^{-kx}, and that hits on various units are independent. Then the probability of an organism losing its viability is $(1-e^{-kx})^n$. If we write u_i for the logarithm of the proportion of organisms not surviving at dose x_i then $E(u_i)=n\log(1-e^{-kx_i})$. The parameters to be estimated are k and n. An iterative method of solving the (non-linear) least squares equations is given. The asymptotic

variance-covariance matrix (under normality assumptions) is given to permit approximate interval estimation. LINCOLN MOSES, *Stanford University*.

Kruskal, William H., "On the uniqueness of the line of organic correlation," *Biometrics*, 9 (1953), 47–58.

For some purposes it may be convenient to represent a multivariate distribution by a single straight line. This paper considers the line passing through the mean of the distribution and having direction numbers proportional in absolute value to the standard deviations, and with their signs determined by the signs of the covariances. This is called the line of organic correlation. It is shown that this is the unique line based on first and second moments which transforms reasonably under omission of coordinates or under change of origin or scale, and which also provides the proper directions of association. If the multivariate distribution is normal then the line is shown to maximize the probability of correct prediction in a certain sense. Certain geometrical properties of the line are proved (no assumption of normality is made). Problems of sampling are not considered. LINCOLN MOSES. *Stanford University*.

Kupperman, M., "On exact grouping corrections to moments and cumulants," *Biometrika*, 39 (1952), 429–34.

Corrections for the cumulants are given for the rectangular and triangular distributions; corrections for the mean and variance are given for the semi-triangular (right-half of the triangular), parabolic, and exponential distributions. R. L. ANDERSON, *North Carolina State College*.

Lancaster, H. O., "Statistical control of counting experiments," *Biometrika*, 39 (1952), 419–22.

Various random experiments were performed to study the adequacy of the χ^2-test consistency of counts from a Poisson distribution with small mean and few counts (2–5). The author concludes that, "χ^2 is likely to remain the method of choice in statistical control of counting, regardless of the size of the sample." R. L. ANDERSON, *North Carolina State College*.

Leslie, P. H., "The estimation of population parameters from data obtained by means of the capture-recapture method. II. The estimation of total numbers," *Biometrika*, 39 (1952), 363–88.

Methods are given for estimating the total numbers in a population under assumptions of constant and varying death-rate and dilution of the population. The death rate is allowed to vary both in time and between different groups of animals. Preliminary analysis of a set of data is given which provides for a test of the absence of dilution and for a method of obtaining approximate estimates from a long chain of samples. L. D. CALVIN, *North Carolina State College*.

Lord, Frederic M., "An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability," *Psychometrika*, 18 (1953), 56–57.

Given the performance of an individual on a series of fallible items which sample a specified ability, what is the best estimate of that individual's "true" ability? The author seeks to construct a metric for measuring the ability underlying a test score that will remain invariant under presumably comparable measures of a given ability. The basic parameters in the estimation model are: h_i measure of item difficulty related to the proportion p_i of examinees who answer the item correctly; c measure of true ability; R_i biserial correlation between answer to item i and true ability of examinees. From these basic parameters, assuming that the distribution functions needed are normal, the author derives expressions for the probability that individual a with ability level c_a will answer item i correctly. Given these theoretical probabilities, the author obtains maximum likelihood estimates for c_a and relates these estimates to the usual type of test score. It is shown that in the special case where all items in a test are of equal difficulty and are equally correlated with the ability measured, the maximum likelihood estimate is a simple function of the usual type of test score. Formulas for the standard error of the maximum likelihood estimates are obtained for conditions of the model. Relationships between these standard errors and the discriminating power of the test at various ability levels are determined and procedures for estimating confidence intervals for the true ability score in terms of the test score are given. For tests composed of equivalent items, the shortest confidence interval for the true score as a function of the test score is obtained for test scores slightly above the halfway point between a chance score and a perfect score. B. J. WINER, *University of North Carolina*.

Maritz, J. S., "Estimation of the correlation coefficient in the case of a bivariate normal population when one of the variables is dichotomized," *Psychometrika*, 18 (1953), 97–110.

Given a normal bivariate population in which one of the variates has been dichotomized and the other variate is continuous but restricted in some way. The biserial correlation coefficient is no longer a consistent estimate of the population correlation ρ. An estimate G defined as $b/(1+b^2)^{1/2}$, where b is the estimated regression coefficient of the continuous variate on the dichotomized variate) has been proposed to handle this latter case. The author has adapted the methods of probit analysis for estimating b for various cases of restriction in the continuous variate. The derivations of these methods are presented. Empirical sampling experiments from normal bivariate populations were carried out to obtain information on the sampling distribution of the coefficient G. Comparisons are made between variances of the probit estimates of the regression coefficients and those obtained from other estimates. The empirical results indicate that G is a more efficient estimate of ρ than is the biserial correlation, even in those cases where both coefficients are consistent estimates of ρ. B. J. WINER, *University of North Carolina.*

McIntyre, G. A., "A method for unbiased selective sampling using ranked sets," *Australian Journal of Agricultural Research*, 3 (1952), 385–90.

A novel method of sampling to estimate the mean value of a characteristic is presented for the case where *measurements* of the characteristic are expensive but it is easy to *rank* a sample with respect to it. For example, it may be easy to rank the plants on a certain area of ground with respect to height, weight, or crop yield, but considerably more expensive actually to measure any of these characteristics. The procedure is to form n independent random samples of size n each (i.e., draw a random sample of n^2 and divide it at random into n subsamples of n each), then get a final sample of n by selecting the largest item from the first subsample, the second largest from the second subsample, and so on down to the smallest from the nth subsample. The precision of an arithmetic mean calculated from this final sample of n is considerably greater than for a simple random sample of n. The ratio of the variances for various population forms and sample sizes ranges from 1.33 (negative exponential distribution, $n=2$) to 3.00 (rectangular distribution, $n=5$). The author suggests $(n+1)/2$ as the typical ratio. The paper also discusses estimation of second and higher population moments, use of a priori knowledge about distribution form, errors of ranking, and clustering of sets to simplify ranking. W. A. WALLIS, *University of Chicago.*

Moore, P. H., "The estimation of the Poisson parameter from a truncated distribution," *Biometrika*, 39 (1952), 247–51.

A counter in a physical problem appeared to stick at certain numbers when counting radioactive particles, e.g., when there were more than r emissions in a given interval. The sample is thus truncated at a certain point, although the number of observations beyond this point is known. The author proposes a simple estimate of the Poisson parameter: $x=\Sigma in_i/\Sigma n_i$ where n_i is the number of intervals with i emissions in the interval. The total number of observations beyond the truncated point is $N-\Sigma n_i$. This estimate is slightly biased (of order $1/N$). The variance of x was also derived. This method of estimating the Poisson parameter was applied to two series of data and found to agree favorably with the maximum likelihood solutions. T. W. HORNER, *North Carolina State College.*

Rudra, A., "Discrimination in time series analysis," *Biometrika*, 39 (1952), 434–39.

A sequential test procedure is presented to decide if a given time-series is of a random, autoregressive, or moving average type. If one of the latter two, it is shown to have the lowest possible order. The test procedure was applied to 28 series and the results compared with the decisions based on other techniques. The probabilities of making various decisions based on 100 operations were computed for two different known structures. R. L. ANDERSON, *North Carolina State College.*

Rushton, S., "On a two-sided sequential t-test," *Biometrika*, 39 (1952), 302–8.

A sequential procedure is given to test the hypothesis that the mean, μ, of a normal population is zero against the alternative that $\mu=\pm\delta\sigma$, where δ is fixed and σ must be estimated. The results are extended to the problem of testing the difference between two means. R. L. ANDERSON, *North Carolina State College.*

Skellam, J. G., "Studies in statistical ecology. I. Spatial pattern," *Biometrika*, 39 (1952), 346–62.

A number of distributions arising in quadrant sampling are considered in relation to the underlying pattern of organisms. It is shown that the same distribution may arise from several quite distinct models. A few ways are briefly suggested, by considering additional evidence of a different kind, as to how to decide whether a given model is appropriate. L. D. CALVIN *North Carolina State College.*

Stevens, W. L., "Samples with the same number in each stratum," *Biometrika*, 39 (1952), 414–17.

Some results are given on the efficiency of constant number vs. proportional sampling. The approximate efficiency is $E=m^2(1-F)/m_2-Fm^2)$, where m is the first moment and m_2 the second moment (about zero) of the frequency distribution of the number of units per stratum, and F is the sampling fraction. R. L. ANDERSON *North Carolina State College.*

Whittle, P., "Tests of fit in time series," *Biometrika*, 39 (1952), 309–18.

A general least squares test of fit of time series models is presented. The statistic is shown to be asymptotically distributed as χ^2; in the limit, the statistic is the ratio of the geometric and arithmetic means of the residual variates' periodogram. The examples included are concerned mostly with autoregressive schemes, but it is emphasized that the tests are appropriate for other methods of graduation. In examples, the test amounts to accepting the hypothesis giving the best fit. D. GOSSLESS *North Carolina State College.*

Williams, E. J., "Use of scores for the analysis of association in contingency tables," *Biometrika*, 39 (1952), 274–89.

Given a contingency table with cell frequencies n_{ij}, $i=1, \cdots, p$ and $j=1, \cdots, q(p\geqq q)$ where $\Sigma_i\Sigma_j n_{ij}=n_{..}$. If fixed scores are available for both classifications, a simple formula is given to estimate the correlation coefficient, r, to be used as a measure of association. The ratio $(n_{..}-2)r^2/(1-r^2)$ is approximately distributed as $F(1, n_{..}-2)$, where the numbers in the parentheses refer to numerator and denominator degrees of freedom of the variance ratio distribution. If only the p row scores are fixed, a method is given to estimate the q column scores, so as to maximise the multiple correlation, R. In this case with ratio $(n_{..}-q)R^2(q-1)(1-R^2)$ is approximately distributed as $F(q-1, n_{..}-q)$. When neither set of scores is fixed, R and the scores are estimated by a canonical analysis. The estimate of R^2 is the largest latent root of a matrix whose elements are simple functions of the cell and marginal frequencies. Tests of significance for a proposed set of scores are presented for $q=2$ and 3. An example is presented with $p=4$ and $q=3$ and 4. Some indications are given in an appendix of the adequacy of the approximations used. R. L. ANDERSON *North Carolina State College.*

Youden, W. J., and Connor, W. S., "The chain block design," *Biometrics*, 9 (1953), 127–40.

Most experimental designs used in agriculture, biology, psychology, etc., involve a fairly high degree of replication, which is appropriate because of the magnitudes of variability encountered. Physical measurements (e.g., spectroscopic determinations) are ordinarily made with far greater precision and a high degree of replication represents a waste of resources. The chain block design is a very elastic arrangement calling for two determinations (each in a different block) on some treatments, and only one on the others. This results in an over all degree of replication which lies between one and two. Methods for layout and analysis are given; practical considerations influencing choice of layout are considered; a numerical example (42 "treatments," of which twelve are repeated, in three blocks) is worked out in detail. LINCOLN MOSES, *Stanford University.*

BOOK REVIEWS

Cyclical Movements in the Balance of Payments. *Tse Chun Chang.* Cambridge (England): Cambridge University Press, 1951. Pp. x, 224. $3.75.

See the article by Solomon Fabricant, pp. 79–87 in this issue.

Demand Analysis: A Study in Econometrics. *Herman Wold in association with Lars Jureén.* New York: John Wiley and Sons; Stockholm: Almquist and Wiksell, 1953. Pp. xvi, 358. $7.00.

See the article by H. S. Houthakker, pp. 88–96 in this issue.

Facts from Figures. *M. J. Moroney.* Baltimore, Maryland: Penguin Books, Inc., 1953. Pp. 472. $0.85.

This constitutes a minor revision in content, but a major downward revision in price, of the volume reviewed by M. A. Girshick in last September's issue of this *Journal* (Vol. 48 (1953), 645–47). The changes which have been made in content are described by the following sentence from the Preface: "The contents remain almost unchanged, except for the latter part of Chaper [*sic*] II which I have revised to include a new approach to modified limit control charts." Changes which have *not* been made are described in the following two sentences: "I am sorry still to remain *persona non grata* to the index number man and the fortune tellers, but there it is. I give way to none in my admiration for the theory (may its shadow never be less!), but when it comes to a great deal of the practice I simply cannot help chuckling."

W.A.W.

The Application of Operations Research to Industry. *Ellis A. Johnson* (Director, Operations Research Office, Johns Hopkins University, Chevy Chase, Maryland). Published by the author, 1953. Pp. 61. Paper. Free of charge.

A. W. Swan, *Courtaulds, Ltd., Coventry, England*

The writer finds this an exasperating publication because parts of it are so extraordinarily good and provocative, while parts seem to wander off into complexities that have little useful interest to the worker in Operations Research.

The introduction is one of the best parts of the book, with a penetrating analysis of the basic thinking in O.R. The author points out that the methods of Operations Research are closer to those of the basic sciences than to those of engineering, but that the techniques and methodology have much in common with those of industrial engineers and management consultants. He goes on to say that O.R. has been concerned from the start with the decision-making system in general, and with the problem of providing indi-

vidual executives with management advice. He considers that one of the main contributions to Operations Research is the use of the team and that it has, more consciously than industrial engineering, developed action-models based on fundamental theory. He also feels that it has relied much more upon complex mathematical concepts and techniques and has realised more fully than industrial engineers the necessity of estimating the uncertainty of its predictions. "Operations Research places a particular demand on the analyst's ability to translate his findings into language which simply and clearly sets forth the values, effectiveness and costs of a set of proposed courses of action."

After this stimulating introduction the author has a chapter, which the writer finds baffling, on the "Relation of the Operations Analyst to the Executive," with a set of diagrams illustrating the interactions of various departments and factors. The ordinary O.R. worker would consider it a waste of time to set these down.

The following sections, "The Operations Checklist for Solving Action Problems" and "Planning Detailed Operations" set out principles in diagrammatic form, and these diagrams are presumably correct, but they are, from the analyst's point of view, what "Punch" calls "glimpses of the obvious," since they are so much taken for granted by the O.R. analyst and industrial engineer that they have become sub-conscious and do not need to be shown as complex diagrams.

Chapter III gives a brief description of "Some Selected Analytical Tools." Unfortunately, statistical method, certainly the most useful and widely applied O.R. tool today, is dismissed in a brief paragraph as being well outlined in Morse and Kimball—an opinion which might not be shared by everyone. Statistical method is not the only new tool used by the O.R. worker to distinguish him from the industrial engineer who preceded him, but it can be stated with some confidence that a very large proportion of present day industrial O.R. work is based upon the statistical approach; the whole gamut of statistical method is used and the statistician is an essential member of any worthwhile Operations Research Department. At this point, any publication on Operations Research must necessarily devote a good deal of attention to the statistical approach and the methods of applying statistical thinking to industrial problems.

The next tool mentioned is Symbolic Logic, but unfortunately the example given appears to be one of very few in which this tool has been applied, and a description of the same example is given in *Factory*, October, 1953. We then proceed to the "Theory of Value" and while this is doubtless a useful method, the writer is not aware of any example in which it has been applied. The following two sections are, "Queueing Theory" and "Stochastic Processes," both of which are, in effect, subsections of statistical method. The "Theory of Games," the next section, is, in the opinion of a large number of O.R. workers, a highly important potential tool, to be used mainly in conjunction with what is known as linear programming, and there is a con-

siderable literature on the subject. The practical use of this kind of thinking has, however, not yet proceeded very far, and this again is a potential rather than actual method.

We then have Chapter IV which gives examples. Unfortunately we start with some "polishing of medals" with two examples taken from wartime practice. The next example is taken from agriculture—not perhaps the most useful for the industrialist. There is also included a brief reference to the massive problem relating to the standard of living in Puerto Rico, interesting, but not of much immediate use to the potential O.R. worker. The author then includes a summary of an excellent paper by John F. Magee of Arthur D. Little, Inc., "The Effect of Promotional Effort on Sales"—the only purely industrial example.

In the final brief chapter on the difficulties met in Operations Research the author returns to his penetrating analysis and the result is excellent. The first difficulty mentioned is that of communication between the scientist and the executive. The O.R. analyst has to learn the executive's language and how to translate into that language. "An operations research study becomes effective in proportion to the amount of effort spent in communicating the effects of the research and clearing up with the executive on a personal basis all the questions involving the validity of the study. Since very few analysts are adept at, or recognise the need for such ability on their part, the results of much good O.R. are never used." The author lists other difficulties in which he includes the extreme difficulty in getting highly skilled specialists from very diverse and often antagonistic disciplines to work well as a closely integrated team. This difficulty can, however, be readily overcome by adopting for every job the simple plan of having the O.R. analyst in charge form a team, consisting of the appropriate members of his own staff and the technicians and other specialists whose knowledge will be most valuable, on the basis that the team will work together for a common aim, and that each member of that team will stand to gain personally in kudos.

In connection with this review it is fair to point out that the subject of Operations Research is a thorny one. There is, today, no completely satisfactory book on the subject and anyone who has the courage to tackle the subject as Ellis Johnson has done deserves praise, especially if he has succeeded in giving useful suggestions, as is certainly the case in this book.

The Revision of the Rapid Transit Fare Structure of the City of New York. *William S. Vickrey.* New York: Mayor's Committee on Management Survey of the City, 1952. Pp. xii, 156. Paper.

WILLIAM R. BUCKLAND, *London Transport Executive*[1]

THIS report is the third of the Technical Monographs from the Finance Project of the Mayor's Committee on Management Survey of the City of

[1] The views expressed are purely those of the reviewer and are not necessarily the views of the organisation.

New York. Its author, one of the Staff Members of the Finance Project, is an Associate Professor of Economics at Columbia University.

After a short introduction on the background of the problem, the report puts forward a marginal costs basis for deciding upon a fare structure to promote the optimum utilization of the particular transit facilities under discussion. This theme is further developed in Chapter 4 (and its Appendix) although it is preceded by a chapter on "Patterns of Traffic"—which is largely an account of trying 'to make bricks without (statistical) straw'—and followed by a development of this traffic theme on the lines of the difficulties of adjusting services to conform to the traffic pattern. The mechanics of fare schemes and collection devices are dealt with in Chapters 6, 8 and 9 while the intervening Chapter 7 again develops the traffic theme in relation to fare changes. Finally there are two short chapters on considerations of equity and what may be called general Social Planning.

The economic theme of the report may perhaps be illustrated by two extracts from pages 4 and 5:

> Since fares must necessarily be set in advance and announced to potential passengers if they are to have the proper effect upon the passenger's decision to travel or not to travel at a given time and place . . . .

> Only if the fare fully reflects at all times and between all points the costs of carrying additional passengers will the fare structure achieve an efficient utilization of the facilities . . . .

The development of the principle ultimately produces a set of proposals on the desirable fare structure for which it would be difficult to carry out the intentions expressed in the first of these extracts, and which would be bewildering to the travelling public if they could.

The important point of the non-monetary costs of travel in the form of fatigue and time costs is brought out very well but the effect is somewhat marred when it is recorded that:

> the passenger would be willing, in order to avoid the inconvenience, to pay an amount that would cover the additional money costs (of providing additional services) . . . .

The idea of passengers being *willing* to pay more money is quixotic since these hedonic cost components are capable of infinite variation as between individuals as well as in time and space. Given a fare structure and a pattern of services, the traveller will tend to minimize his total costs—monetary and non-monetary—according to the needs of the moment and changes in these needs are, at this level of detail, part of the random fluctuations which are present in all patterns of travel. Therefore it is suggested that the fare structure which will help the traveller most in this task is one for which the money cost is virtually the same for comparable distances at any time of the day. In this way the passenger has the greatest opportunity to work out his own salvation and the traffic flows unconstrained by differential fares. It must be said, however, that this kind of scheme requires a degree of assimilation of

fare structures for all the major forms of public passenger transport which may not at present be available in New York. It seems to this reviewer that the disadvantage with schemes for differential fares according to time and place—such as are largely advocated in this report—is that as soon as a fare structure is promulgated its operation tends to change the flows of traffic upon which it is based and it then becomes necessary to change the fare structure if the optimum position is to be maintained. Thus, the fare to be paid for a given journey is a matter of some doubt for the travelling public which must be psychologically undesirable.

To deal with some rather more practical issues: on page 10 there is an equation expressing operating expenses in terms of certain physical operating characteristics. The equation appears to be the usual form of multiple regression equation assuming no interaction between the various characteristics but it has, in fact, been developed by the non-statistical process of allocating the operating expenses to the various physical characteristics of operation. For example, the salaries of motormen are allocated to the characteristic of train miles. This may have been the only reasonable method available but it would have been desirable to have given more space to the interpretation in order to avoid possible confusion. In Chapter 3 there does not appear to be any reference to the survey work on travel in New York which has been done by various organizations in connection with transportation advertising and which should have been useful for this investigation. In connection with the availability of data for this study it would have been useful to have put forward some recommendations on how the various deficiencies might be filled. The detail in Chapter 9 of the various collection devices associated with the different kinds of fare structures discussed in this report give the reviewer a distinctly unhappy feeling as to their practicability both from the engineering and the commercial point of view. There appear to be far too many things capable of going wrong. These vary from the station staff not changing something vital at the right time, through the peak-period passenger not having ready the right coins, to the mechanical and electrical devices necessary to display illuminated signs for the currently required fare (with gongs to signal the alterations) and the amount of change the passenger may expect to receive as a result of his not having the correct coins available.

The considerations of equity and social planning at the end of the report begin to place the whole subject into a perspective in which passenger transport takes on the aspect of something which is very much concerned with the human business of living, working, and playing. Transport is only a means to an end and that end may well be the optimum utilization of the human and material resources of a given area, say New York. Surely, since this investigation was worthwhile, it should have been approached in the spirit of an operations research project, for its solution demands the welding of economics to transport operation and engineering with statistical methods as the flux. As it is, an undue concentration of attack on the first of

these has produced a report which will probably make the transport operator and his engineering colleagues shudder: the statistician will merely lament once more that most of the data necessary for the problem were not available.

Measuring Your Public Relations, a guide to research problems, methods and findings. *Herman D. Stein.* New York: National Publicity Council for Health and Welfare Services, Inc., 1952. Pp. 48. $1.25.

MARIE JAHODA, *New York University*

THE aim of this booklet is to provide the professional personnel in health and welfare organizations with a balanced view of what research can do for their agencies and to describe different research procedures adequate for different types of problems. Mr. Stein does not want to "sell" research; rather he wishes to enable persons concerned with the practical work of health and welfare agencies to decide for themselves what research they need.

With this aim in mind he discusses: the nature of the problems which arise in the public relations of voluntary agencies; informal research techniques (they might better be called fact-finding techniques) which an agency can apply largely without expert help; pre-testing of written material; the value and limitations of public opinion polls; communications research, etc.

The presentation of these matters is straightforward without being oversimple. There are many examples used in the text to illustrate a point. A small list of selected references on a more technical level completes the publication.

It lies in the nature of such a publication that it offers no new ideas. The research person who is about to embark for the first time on a study for or of an action agency might find it helpful, however, to glance through this booklet. Whether or not it fulfills its aim with its actual target audience is in itself a question for research in public relations.

The WOI-TV Audience. Mimeo Series No. 1. Ames, Iowa: Statistical Laboratory, Iowa State College, 1952. Pp. 125. Paper.

LESTER R. FRANKEL, *Alfred Politz Research, Inc.*

The WOI-TV Audience is a statistical report describing the size and characteristics of the television audience located within a 50 mile radius of Ames, Iowa. The data are based upon a sample survey, and were obtained for the purpose of establishing bench mark data against which future surveys may be compared. The text material in this report is a description of the methods used to obtain the data.

The survey design, the questionnaire, the sample plan, and the field operations were the responsibility of the Statistical Laboratory, Iowa State Col-

lege and, as is to be expected, the techniques used appear to be superior to those employed in the usual stereotype commercial television audience study.

The sampling procedure was designed in such a manner as to make full use of the resources that were available. The sample was of the multiphase, single-stage type. After a sample of households had been selected, household characteristics data relating to the ownership of TV sets was obtained. In the second phase, additional data were obtained from all television households, 25 per cent of the non-television households and 50 per cent of all the adults in these two groups of households.

Single-stage sampling was employed in the selection of the households to be included in the sample. A sample of 400 segments was selected to represent the survey area, and all households within each of the designated segments (approximately 6) were included in the first phase sample. Of particular interest to the practicing sampling statistician is the discussion of the uses of the Master Sample Maps, the City Directory, photocounting, and the methods of cruising for the determination of segment sizes.

Aside from the problem of sample design (which is merely a blueprint) the problems of the execution of the survey are discussed in detail. The training was particularly important in view of the fact that the second phase of sampling was accomplished by the interviewers at the field level. In addition, since some of the questions on the questionnaire dealt with the respondent's activity on the day before the interview, it was necessary to spread the interviewing equally over all seven days of the week.

The format and organization of the report as well as the presence of some minor arithmetical inconsistencies tend to detract from the impressiveness of the study. However, it is clear from the description that the design was not intended to produce a quality impression but a quality study in the most efficient manner possible. For example, ten years ago it was more or less assumed that the steps in selecting a sample should follow the time consuming sequence of primary unit selection, segment selection, listing of addresses by interviewers, final household selection in the central office from the often shaky prelistings, and interviewers revisiting the selected locations to interview. In the study of the WOI-TV audience there was an efficient utilization of man hours. Real inventiveness based on genuine statistical knowledge obviously played a role in finding an additional method by which the household and individual selection was accomplished in a single field operation.

Social and Psychological Factors Affecting Fertility. Volume Three. *P. K Whelpton and Clyde V. Kiser.* New York: Milbank Memorial Fund, 1952. $1.00. Paper.

E. Lewis-Faning, *Welsh National School of Medicine*

FOURTEEN years ago, Raymond Pearl (1939) in his book "The Natural History of Population" calculated pregnancy and live birth rates for

various groups of women, differentiating those who had, and those who had not used contraceptive methods. He discussed among other things the effects of contraceptive efforts on the pregnancy rates in relation to economic status, education and religion.

Ten years later, the Royal Commission on Population (1949) compared pregnancy rates for periods of reproduction in which (*a*) no birth control was used, (*b*) contraception was abandoned, (*c*) contraception was being used. They also compared the average desired and actual size of the family, and the number of unwanted children for groups of women classified according to the degree of success attained in planning and spacing a family by contraceptive methods.

The collection of seven papers here reviewed goes still more deeply into the subject and inquires what social and psychological factors contribute towards successful family planning. Is it those who feel most economically insecure who successfully restrict the size of their families? Is it those who plan other aspects of their lives? Is it those with poor health or those with a feeling of personal inadequacy? These are among the problems which constitute the subject matter of these papers.

It will be noted that in both the earlier publications the indices used were quantitative—or, if not, were definitely factual—grade of education reached or religious denomination—and that to such data statistical methods and reasoning could legitimately be applied. The reader's reactions to the volume under review must depend on whether or not he accepts that indices of such nebulous qualities as "a feeling of economic insecurity" or "a tendency to plan in general" or "a feeling of personal inadequacy"—indices to which statistical methods and reasoning have here also been applied—have been or indeed can be satisfactorily constructed.

One example will illustrate the dubiety which the reader should not fail to feel. Replies by 1444 wives to the question "Do you plan your buying to take advantage of sales?" were tabulated as below, showing the distribution according to success in family planning ("A" being the most successful) of the group of women giving each specific answer.

On this, the authors comment: "One group of wives answered "Very

Plan to buy at sales	No.	Percentage distribution by fertility planning status			
		A	B	C	D
Very often	479	31	13	31	24
Often	481	25	17	30	28
Sometimes	419	28	13	35	24
Seldom	36	31	11	19	39
Very seldom	29	21	7	21	52

Often." Among these, 44% were in the effective fertility planning categories (A and B). Only 28% of the wives answering "Very seldom" were in the effective fertility-planning groups."

Can answers such as these be used as indices of "a tendency to plan in general"? If one has been reared in the belief that most of the goods offered in sales are manufactured especially for sales and are generally of an inferior quality, one plans *not* to buy at sales. Nevertheless, if there is a large family, the wife may be forced to buy in the cheapest market. Does this make her a (good) planner? In fairness, it must be stressed that in the instance cited, the conclusions are based on the replies to many more questions of the same type than the one used as illustration. The volume as a whole must contain some hundreds of tables like the one quoted. There is no lack of data and the arguments from interpretation of the figures, although involved, seem reasonable, always provided that the indices used do really measure what they are supposed to be measuring. On this fundamental matter some scepticism is not unjustified.

In one section, scores are allocated to sets of replies to different questions, and Pearsonian correlation coefficients calculated as a measure of the intercorrelation between the different indices used, in spite of there being no evidence that the intervals represented by differences between the scores were uniform. Is the difference between "Very often" and "Often" equivalent to the difference between "Often" and "Sometimes"?

Qualified statisticians reading these papers can safely be left to utilize their professional knowledge and experience in assessing the statistical validity of the conclusions reached but there is a real danger that the statistically unsophisticated (who comprise possibly the majority of workers in this field) will be overawed by the facade of diagrams and statistical tables into accepting the conclusions as authoritative and final. To such, two things need pointing out: first, that the 1444 couples whose reproductive and contraceptive histories and personal opinions as to the impact of their economic circumstances and psychological characteristics on their desire for children, form the data common to all seven papers, although a homogeneous group and therefore presenting certain advantages for studies of this type, are far from being representative of any past or present community in the U.S.A.; second, that the answers recorded on the questionnaires were given as long ago as 1941–42 by couples married in the years 1927–29 and reared in the traditions of 1900–10. Strictly, then, the replies should be interpreted in the light of the economic circumstances that prevailed when they were building their families—mainly in the era of depression of the late 20's and early 30's—rather than in the light of those of the 1950's. That the authors have records enabling them to do this, or indeed that they realize the essentiality of attempting it, is not clear. Many aspects of life changed even between 1941 and 1951.

Statisticians, especially those interested in the field of human fertility, would be rendering a service by reading and commenting on this publication.

REFERENCES

Raymond Pearl, *The Natural History of Population*, Oxford University Press, London, 1939.

Papers of the Royal Commission on Population, Vol. 1, *Family Limitation and its Influence on Human Fertility*, H.M. Stationery Office, London, 1949.

Community Wage Patterns. *Frank C. Pierson.* Berkeley and Los Angeles: University of California Press, 1953. Pp. xvii, 213. $3.75.

MITCHELL O. LOCKS, *University of Oklahoma*

THIS volume is an attempt to determine the nature and causes of relationships between post-World War I wage developments in Los Angeles County, California and those in other large metropolitan areas. The book has chapters on each of the following topics: Pre-1940 Wage Levels; 1940–1949 Wage Levels; Local Influences on Wage Levels; Industry Influences on Wage Levels; Relationship Between Employment and Wages; Relationships Between Investment, Productivity, and Wages; and The Influence of Unions on Local Wages. As can be seen from the titles, each chapter covers a topic so important that it could be a separate study in itself.

The author has used a great wealth of material assembled from many different sources, principally other books about California. However, with a study having such broad scope confined to only 164 printed pages (besides Appendix Tables), it is not surprising that at times the author seems to wander aimlessly in a forest of secondary statistics. The reviewer received the impression that the author was not sure of what his data showed him, and therefore had to state many of his conclusions without strong conviction.

An example of this may be found in his analysis of the relationship between Employment and Wages. He computed certain rank correlation coefficients between employment and wages for the period from 1929 to 1939, and also for the period from 1940 to 1949. His average rank correlation coefficient for six cities for the earlier period was statistically significant (although the individual rank correlation coefficients for each of the six different cities for that period were not), while the average rank correlation coefficient (as well as the six individual rank correlation coefficients) for the later period were not significant. On the basis of this very meagre evidence, the author makes the following statement without further explanation:

> This suggests that industry employment and wage levels move together when there is a large amount of unemployment, but that these two variables bar no consistent relationship to each other when labor market conditions are relatively tight. (p. 116)

This statement contradicts the widely-held view that the elasticity of labor supply with respect to wage rates is high in depression. For that reason, the author missed an opportunity to make a significant contribution to wage theory by not further explaining the reasons for this observation.

Another example of inadequate explanation occurs on page 61:

> . . . hourly earnings in Los Angeles manufacturing rose less rapidly than in the United States between 1945 and 1949 (30 per cent as against 37 per cent).

However, data obtained from Appendix Table 1 show that Los Angeles retained its ranking in average manufacturing wage levels during the period from 1945 to 1948. (Comments concerning the validity of some of the data in Appendix Table I will be found later in this review.) These data purport to show that Los Angeles was seventh in average manufacturing industry wage levels in a group of 20 large cities in both April, 1945 and April, 1948. Although this discrepancy in findings made at two different points in the book is a relatively minor one, the author should have furnished an explanatory note.

The overall statistical facets of this book inspire the following comments:

(1) *Applications of the Analysis of Variance.* In Chapters VI and VII the author performed analyses of variance on eleven different manufacturing industries in six different cities with respect to percentage changes in employment, average annual earnings per worker, and value added per worker from 1929 to 1939. Separate analyses were performed for each of these three characteristics using the "F" test, and the results were reported in Tables 11 and 20. However, examination of Tables 8 and 19 show that the same data were ranked within communities for purposes of computing rank correlation coefficients. Thus the data in those tables were already in a form such that with a negligible amount of additional calculation, a "Friedman" rank analysis of variance could have been performed.[1] The reviewer submits that the latter method would be preferable because of the normality assumption implicit in the "F" test.

(2) *Appendix Table I.* There is reason to doubt the validity of the data in the last column of Appendix Table I. Since much of the analysis of wage movements in Chapters II and III is based on that table, there should be further explanation of how those data were obtained. The reasons are:

(*a*) The source given for the data in that table is a book published in 1946.[2] Yet this fourth column gives manufacturing wage indices for April, 1948.

(*b*) The data in the first three columns of that table give, respectively, manufacturing wage indices for 20 different cities for April, 1941; April, 1943;

---

[1] Milton Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, 32 (1937), 675–701.

[2] Ruth McFarlane. *Wage Rate Differentials: Comparative Data for Los Angeles and Other Urban Areas.* (Los Angeles, 1946.) The Haynes Foundation.

and April, 1945, each adjusted to an average index of 100 for its date. However, column 4 has an average index of 144 for the 20 cities. Thus it would appear that a different basis was used for calculation of the April, 1948 indices than had been used for the earlier three dates.

The reviewer believes that despite indicated shortcomings of the book, the author made a contribution to the study of wage patterns. In large measure, the scope of his analysis is limited by the fact that he had to use secondary data. However, there is need for more studies of this type using more primary data than are now available for this purpose.

Punch-card Methods. *Harry P. Hartkemeier.* Dubuque, Iowa: Wm. C. Brown Company, 1952. Pp. xvii, 360. $5.00. Paper.

P. C. Hammer, *University of Wisconsin*

The subtitle of this book is "How to Use and Operate Punching, Sorting, Electronic Statistical, Tabulating, and Accounting Machines Including Types 24, 26, 75, 80, 82, 101, 402, 403, and 407." All the machines discussed are the IBM models. Basically the book is an illustrated reassembly of materials contained in the manuals provided by the IBM Corporation free of charge. For purposes of instruction this arrangement may be better than the individual manuals.

The book is slanted toward commerce and accounting students, the problems being primarily in those fields of interest. The most difficult mathematical problem dealt with seems to be progressive digiting. Since there is a decided lack of good expository material on punch card methods more books in the field are indicated. However, the reviewer feels that the author has neglected some of the basic punch-card equipment and methods in this book.

For example, the summary punches, the collators, reproducers, and the calculating punches (602A and 604) are not discussed although each is of great usefulness in statistical and commercial work. The author gives the impression that all machines not discussed are no longer being manufactured. This is by no means the case; all the additional machines mentioned above are still in production.

Since the book neglects so many virtually essential machines for scientific and accounting practice it cannot be recommended as a text without extensive supplementary materials.

The book is reproduced by photo-offset and has a soft paper cover and a permanent loose-leaf binding. The text is well written in view of the necessarily segmented character of such manuals.

Associated Measurements. *M. H. Quenouille.* New York: Academic Press Inc., 1952. Pp. x, 242.

ISADORE BLUMEN, *Cornell University*

HERE is a handbook on correlation, regression, and related topics which many statisticians will find useful. The formal layout of analyses, hints on practical pitfalls, and the details of manipulative technique are illustrated through the copious use of numerical examples.

Unfortunately, the very nature of such a handbook seems to have forced the omission of so much in the way of basic ideas that it cannot be recommended to non-specialists. The statistician will want the book in order to have details readily available. The more general reader, however, will be bothered by such fundamental problems as the rationale for the choice between various methods proposed. For him there are not always clear answers.

The book is divided into four parts. The first forty-eight pages contain those "quick and dirty" methods which the author finds most useful and which require only plotted data. A section on similar numerical methods is included later under the heading of grouping observations. Included in these sections are graphical methods for bivariate and multiple correlation problems and curvilinear regression, a few non-parametric tests, and some devices adapted to situations where data is easy to obtain, great accuracy not wanted, and simple computation important. Omissions, due apparently both to the organization of the book and to the desire to keep it down to manageable size, include most non-parametric procedures—e.g., the rank correlation coefficient (which name the author chooses to bestow on Kendall's tau) and runs tests. Biserial, tetrachoric, and related correlation procedures are not mentioned.

The second part covers the conventional topics: bivariate correlation and regression, multiple and partial correlation, and curvilinearity. This is well done, although one might quibble that the author's treatment of such problems as identifiability of parameters is not strictly accurate, the conditions he gives being sufficient rather than necessary. The problem of using correlated variates for screening and selection, as in personnel testing and genetics, is not treated.

The third part includes grouping, analysis of covariance, and general pointers on the organization of investigations. Readers who have not been exposed to covariance before should be warned that this exposition is not particularly lucid.

The last sixty or so pages deal with a variety of problems. The section on multivariate analysis is remarkably well done for so condensed a non-mathematical discussion. The problems of time series, not being easily reducible to elementary terms, are somewhat less satisfactorily treated but this section will nevertheless be quite useful. There is also a section devoted to a variety of hints and comments.

Most of the more desirable tables are included as is a fairly extensive,

but not selected, bibliography. Teachers who use this as a laboratory manual in their courses, for which the book is well adapted, may complain of a lack of problems for students to work out for themselves. It would also be desirable if in future editions the conventional names for tests and procedures were used.

From an over-all point of view, the reviewer was disturbed by the lack of discussion of alternate hypotheses, of the power of various tests proposed and of the relative quality of various estimation procedures. Why reject extreme observations in one case and not in another? Why use a graphical method instead of the more common estimates? Why choose one nonparametric device over another? Surely more thoughtful answers could have been provided by so competent an author.

Hypothesis Testing in Time Series Analysis. *Peter Whittle.* New York: Hafner Publishing Company, 1951. Pp. 120. $3.50. Paper.

JOHN GURLAND, *Iowa State College*

IN THE subject of time series analysis, where the paucity of suitable statistical tests is conspicuous, a book of this sort is a welcome addition to the literature. It abounds in suggestions and ideas which should stimulate more research in this area.

The spirit of the book is commendable. It attempts to give a general rationale for discriminating between different random structures which might be regarded as having generated the same observed time series. The null hypothesis and the alternatives are always stated explicitly. Then tests are constructed which presumably are optimal in some sense. Some ingenious ideas and devices are propounded but Whittle is somewhat carried away in his enthusiasm, with the result that clarity and rigor are sometimes sacrificed for expediency. The reader of this book should be warned that this is not a book which may be read uncritically, but rather one which should be read with caution and reserve.

The first two chapters comprise a brief review of some important results in statistics and probability theory, with a few sketchy proofs. Chapter 1 outlines the testing of hypotheses and the construction of a most powerful critical region by means of a sufficient estimator. The second chapter reviews the notion of a spectrum for a stationary stochastic process and gives the corresponding spectral expansion of the process. The discussion centers mainly on a discrete process as this is the type considered throughout the book. By restricting the spectral density to be a rational function of $z=e^{i\theta}$, it is shown that the corresponding stochastic process is either an autoregressive scheme or a moving average scheme or a certain generalization of these.

In Chapter 3 a most powerful test is constructed, on the assumption of an underlying normal distribution, for testing whether N consecutive obser-

vations have a particular covariance matrix. The test criterion is a ratio of quadratic forms, and is the same as that given by Lehmann and Stein (*Annals of Mathematical Statistics*, 1948, p. 504). If the covariance matrix is assumed to be a Laurent matrix (as is the case for a stationary process), and N is large, a test function is constructed which is a ratio of linear functions of the empirical covariances, and which is simpler to compute than the aforementioned ratio of quadratic forms. The large sample distribution of the test is given and Whittle states that this test in the case of a stationary process has "practically the same power" as the exact test mentioned above. The reviewer cannot resist wondering what happens in the case of small or moderate values of N, since both the construction of the test and the distribution of the criterion assumed large values of N.

Chapter 4 gives some ingenious approximative methods for getting the inverse of a Laurent matrix, also its latent roots. Circulant matrices are used in the approximation and the spectral density of the process is elegantly applied. It is not at all clear, however, how good are the approximations. As for the approximate distributions of quadratic forms and ratios of quadratic forms given in this chapter, the reviewer would like to make a few comments.

In the case of a quadratic form Whittle's proposal of employing the Edgeworth form of the Gram-Charlier series requires investigation. So far as this reviewer is aware there is no published theory which established the validity of this series as an asymptotic expansion for the distribution of a quadratic form in correlated random variables. P. L. Hsu (*Annals of Mathematical Statistics*, 1945, 1946) proves some theorems for certain special quadratic forms of independent random variables which justify such asymptotic expansion in these cases. For the case of normally distributed variables the reviewer has some papers in the process of publication which show how a Laguerrian expansion may be used so as to actually converge to the distribution function of a quadratic form.

In regard to a ratio of quadratic forms in correlated normal variables, Whittle recommends finding the moments, then using a Gram-Charlier expansion to approximate the distribution. This method apparently works well in the numerical example given in Chapter 5, but as a general method it has some inherent difficulties which might be pointed out here. In the first place, the problem of finding the moments is, in general, prohibitive. For the special cases considered by Whittle the distribution is required for independent variables, and the denominator is such that the ratio is distributed independently of the denominator. These circumstances greatly simplify the problem. Ordinarily, however, the integrals which represent the moments are of hyperelliptic type. In the second place, the validity of the asymptotic expansion is suspect. If the range of the ratio is finite then a condition of Cramér's is satisfied which assures that the Gram-Charlier series converges to the distribution function. If the range is not finite, then the same questions as above regarding the convergence and the validity of the asymptotic expansion apply.

Some further remarks concerning this chapter are apropos regarding the assumption of circularity. A rather sweeping statement appears in referring to R. L. Anderson's distribution, to the effect that the assumption of circularity is "really no great drawback as N would have to be quite small before the power would be seriously diminished by this assumption." What is meant by the terms "really no great drawback," "quite small," "seriously diminished," is vague here. If N must not be "quite small" for the assumption of circularity to be "seriously" questioned, then one could ask whether or not the normal distribution is an adequate approximation. If, besides the circularity assumption, the characteristic function is "smoothed," as suggested by Koopmans (*Annals of Mathematical Statistics*, 1942) or more generally as extended by Whittle, then one may well wonder how far astray the resulting approximation is from the original distribution before circularity and smoothing were applied. How large or small N must be and what the corresponding effect will be on the power and on the true distribution is indeed a moot question and one which, in the present stage of development of the theory is usually answered by conjectures which to this reviewer seem unduly optimistic.

Chapter 5 provides a numerical example for a test of randomness against certain alternative hypotheses. The approximative methods developed in the earlier chapters are used, and seem to work quite well for this example.

Chapters 6 and 7 are entitled "Non-parametric Discrimination" and are devoted mainly to the problem of constructing suitable tests regarding the structure of a process. The title of these chapters is misleading because the tests are constructed from a probability density involving unknown parameters and, as such, are parametric tests in the conventional sense of the term. Whittle, in fact, assumes the parameters have a probability distribution and proceeds to apply Bayes' theorem to construct a posteriori likelihood functions, then uses a likelihood ratio of such functions. It is surprising that such an anachronistic approach could have found its way in so recent and otherwise modern a book. By assuming that N is large and choosing a convenient distribution for the parameters various test criteria are obtained. Many of the tests could be constructed directly without appealing to Bayes' theorem. Among the tests considered are the following: (a) Test the order of a moving average scheme against the alternative of a different order. (b) Test the order of an autoregressive scheme against the alternative of a different order. (c) Test whether a process is an autoregressive scheme of a fixed order against the alternative of a moving average scheme of a fixed order. (d) Opposite of (c).

In Chapter 8 numerical examples of (c) and (d) are discussed and in Chapters 9 and 10 the methods of the earlier chapters are applied to construct periodogram tests and tests of fit, respectively.

The final chapter "Indeterminacies in model structures" provides an interesting investigation into the non-uniqueness of the linear structure of a stochastic process for a given covariance function. In the case of a Gaussian

process, the indeterminacy is inherent; however, if the process has some non-zero cumulants of higher order than the second, a method of discrimination is proposed.

In conclusion, this reviewer would like to quote from F. N. David's review of this book which appeared in *Biometrika*, Vol. 39, May, 1952. " . . . However it is more than sufficient to say that Mr. Whittle is a pioneer and it has always been the fate of pioneers both to stimulato those who follow and to be criticized by those who are wise after the event. All who are interested in time series will benefit by reading the book if only from the stimulation and excitement which come from trying to go one better than the author. This is an important contribution to the research work on time series and may well prove to be the foundation stone of a satisfactory theory."

Tables of Poisson Distribution. *Tosio Kitagawa.* Tokyo, Japan: Baifukan, 1952. Pp. xii, 156. $3.50.

William G. Cochran, *Johns Hopkins University*

These tables give the individual terms $e^{-m}m^x/x!$ of the Poisson distribution. Unlike E. C. Molina's tables (*Poisson's Exponential Binomial Limit*, D. Van Nostrand, New York, Fifth printing, 1949), they do not contain cumulative sums, and they stop at $m=10$, whereas Molina goes up to $m=100$. However, the interval of tabulation is much smaller than Molina's, being only 0.001 in m up to $m=1$, and thereafter 0.01 up to $m=10$. The following table compares the intervals available and the number of decimal places (D.P.) given by each author.

Range of m	*Tabulation interval*		*Decimal places*	
	Kitagawa	*Molina*	*Kitagawa*	*Molina*
0.001– 0.010	0.001	0.001	8	7
0.010– 0.300	0.001	0.01	8	7
0.300– 1.000	0.001	0.1	8	6
1.00 – 5.00	0.01	0.1	8	6
5.00 –10.00	0.01	0.1	7	6
10.0 –15.0	none	0.1	–	6
15 –100	none	1.0	–	6

For anyone engaged in accurate computations with small values of m, Kitagawa's tables are a valuable addition to the library. They are attractively printed, with ample separation of the figures so as to diminish eyestrain and copying mistakes. The text is in English.

Additional tables give single and double inspection plans. These plans are based on the same principle as those by Paul Peach (*Industrial Statistics and Quality Control*, Edwards and Broughton Co., Raleigh, 1947), but were developed independently by Kitagawa. If α = consumer's risk, β = producer's risk, Kitagawa gives tables for finding the sample size (or sizes) and the rejection number (or numbers) for $\alpha=0.1$, $\beta=0.1$; $\alpha=0.1$, $\beta=0.01$; $\alpha=0.01$, $\beta=0.01$, whereas Peach's tables have $\alpha=\beta=0.05$.

Stechert-Hafner, Inc., 31 East 10th Street, New York 3, inform me that they hope to have a supply of Kitagawa's tables at $3.50 each. From my correspondence with the Baifukan Company, it appears that the company does not wish to promote direct sales from Japan.

50–100 Binomial Tables. *Harry G. Romig* (Quality Manager, Hughes Aircraft Company, Culver City, California). John Wiley & Sons, Inc., 1953. Pp. xxvii, 172. $4.00.

These tables show to six decimal places the individual and cumulative terms of the binomial distribution for probabilities from 0.01 to 0.50 in steps of 0.01 (from which, of course, values from 0.50 to 1.00 in steps of 0.01 are readily obtained) and for sample sizes from 50 to 100 in steps of 5. The introduction defines the binomial distribution, discusses its relation to the hypergeometric and Incomplete Beta-function, explains the notation used in the tables, describes the procedures used in computing the tables and their accuracy, and gives directions for using the tables and for interpolating into them, together with examples.

The Government Printing Office has recently printed a far more extensive table of the binomial distribution—giving, however, only cumulative probabilities—with entries to seven decimals for the same probabilities covered in Romig's table and for sample sizes from 1 to 150 inclusive, by steps of 1. Apparently this table (Ordnance Pamphlet ORDP 20-1, *Tables of the Cumulative Binomial Probabilities*, September 1952) is to be made available to the public, in which case a more definite notice will be included in this Review Section.

W.A.W.

Confidence Limits Tables for Samples of Binomially Distributed Data. *John Folger* (Chief, Technical Services Division, Human Resources Institute, Maxwell Air Force Base, Alabama). Maxwell Air Force Base, Alabama: Human Resources Institute, May 1953. Pp. 12.

These tables give 95 per cent confidence intervals for sample sizes from 5 through 49 by steps of 1, and for all possible numbers of successes. "These

confidence limits tables were prepared from *Tables of Binomial Probability Distribution*, National Bureau of Standards, Applied Mathematics Series 6." No exact definition of the confidence intervals is given, nor any further account of the method of computing. Presumably, however, the intervals are such that not less than 2.5 per cent lies in each tail.

W.A.W.

Cambridge Elementary Statistical Tables. *D. V. Lindley and J. C. P. Miller.* Cambridge (England): Cambridge University Press, 1953. Pp. 35. $1.00. Paper bound.

As STATED in the Preface, "This set of tables is concerned only with the commoner and more familiar and elementary of the many statistical functions and tests of significance now available."

Table 1 shows cumulative normal probabilities to 5 decimals for arguments 0(0.01)3.0(0.1)4 and for all arguments above 3.731, and a brief tabulation of the normal frequency function. Table 2 gives the one-tail percentage points of the normal distribution function for selected percentages. Tables 3, 5, and 7 give percentage points of the t-, χ^2-, and F-distributions.

Table 4 gives the normalizing transformation for correlation coefficients. Table 6 gives a means of estimating the standard deviation of a normal population from the range of a small sample (13 or less). Table 8 gives 4,000 random digits. Table 9 gives the square, square root, reciprocal, reciprocal square root, and common logarithm and antilogarithm of each integer to 1000; it also gives inverse circular and hyperbolic root-sine transformations. Table 10 gives logarithms of factorials to the base 10.

It is the hope of the authors that "the values provided will meet the majority of the needs of many users of statistical methods in scientific research, technology, and industry in a compact and handy form," and that they will be convenient for the teaching and study of statistics in schools and universities.

M.A.L.

County and City Data Book, 1952. A Statistical Abstract Supplement. *Prepared under the direction of Morris B. Ullman* (Chief, Statistical Reports Section, Bureau of the Census). Washington: United States Government Printing Office 1953. Pp. xxx, 608. $4.25.

ACCORDING to the Introduction, "This volume is one of a series of supplements to the *Statistical Abstract of the United States*, and is designed to meet the need for summary statistics for small geographic areas. Compactly assembled in this volume are 128 items of data for each county, standard metropolitan area, State, and geographic division; and 133 items of data for each of 484 cities having 25,000 or more inhabitants in 1950. Also included is a table showing the number of inhabitants of all urban places

(mostly incorporated places of 2,500 inhabitants or more) in 1950. . . . The year, 1952, used to designate this edition denotes the year during which compilation of the statistics occurred." The title page notes: "Statistics included: For 1950, Agriculture, Area and Population, Banking, City Government Finances and Employment, Construction, Education, Family Income, Housing, Labor Force, Vital Statistics, and other subjects; for 1947 and 1950, Manufactures; for 1948, Trade and Services; and Climate."

W.A.W.

Bibliographie sur la méthode statistique et ses applications. *G. Darmois and E. Morice, editors.* Paris: Institut International de Statistique and Institut National de la Statistique et des Etudes Economiques, 1952. Pp. 49. Paper bound.

THIS bibliography lists 75 works dealing with statistical method and its applications which have been written in or, in two instances, translated into French, the majority within the last twenty years. The introduction apologizes to authors whose works may not have been cited, explaining that the bibliography could not be exhaustive.

The bibliography has been divided into two sections; (1) General Methods, and (2) Applications. Under the first heading are (*a*) elementary works, (*b*) intermediate works, (*c*) advanced works on theory, (*d*) elementary probability, and (*e*) probability theory. The second section includes (*f*) economics and insurance, (*g*) industry and agriculture, (*h*) demography, (*i*) medicine, biology, and psychology, and (*j*) mechanics and astronomy.

Besides the usual bibliographical details of author, publisher, date of publication, number of pages, and price, this bibliography gives the table of contents of each work, with a summary of what is included in each part.

M.A.L.

PUBLICATIONS RECEIVED

Allen, R. G. D., and Ely, Edward, Eds. *International Trade Statistics.* New York: John Wiley and Sons, 1953. $7.50.

American Society for Quality Control. *Definitions and Symbols for Control Charts.* New York: ASQC (70 East 45th St.), 1953. Paper. 50 cents.

Anderson, Oscar Edward, Jr. *Refrigeration in America.* Published by the Princeton University Press for the University of Cincinnati, 1953. $6.00.

Bross, Irwin D. J. *Design for Decision.* New York: The Macmillan Co., 1953. $4.25.

Bureau of the Census. *Farms and Farm People.* Washington, D. C.: U. S. Government Printing Office, 1952. Paper. 50 cents.

Burns, Arthur F. *Business Cycle Research and the Needs of Our Times.* 33rd Annual Report. New York: National Bureau of Economic Research, 1953. Paper.

Canada, Department of Trade and Commerce. *Private and Public Investment in Canada; Outlook 1953.* Ottawa: 1953. Paper.

———. *Supply of Building Materials in Canada; Outlook 1953.* Ottawa: 1953. Paper.

Charnes, A., Cooper, W. W., and Henderson, A. *An Introduction to Linear Programming.* New York: John Wiley and Sons, 1953. Paper.

The Cincinnati Report, 1952; An Appraisal by 600 Citizens of Government and Voluntary Services in Cincinnati and Hamilton County, Ohio. 1950–1952. Cincinnati: Council of Social Agencies, 1952. Spiralbound.

Committee for Economic Development. *Flexible Monetary Policy; What it is and how it works.* New York: March 1953. Paper.

Connolly, T. G., and Sluckin, W. *Statistics for the Social Sciences.* New York: Hafner Publishing Co., 1953. $2.75.

England and Wales, The Registrar General's Statistical Review of, for the Year 1951. London: H.M. Stationery Office, 1953. Paper. 5 s.

Ferber, Robert. *The Railroad Shippers' Forecasts.* Urbana: University of Illinois, 1953. Paper. $1.00.

Festinger, Leon, and Katz, Daniel, eds. *Research Methods in Behavioral Sciences.* New York: The Dryden Press, 1953. $5.90.

Gilbert, Milton, ed. *Income and Wealth, Series III.* Cambridge, England: Bowes and Bowes, 1953. 35 s.

Hansen, Alvin H., and Clemence, Richard V., eds. *Readings in Business Cycles and National Income.* New York: W. W. Norton and Co., 1953. $5.25.

Hansen, Morris H., Hurwitz, William N., and Madow, William G. *Sample Survey Methods and Theory.* New York: John Wiley and Sons, 1953. Vol. I, $8.00; Vol. II, $7.00.

Hertz, David B., ed., and Rubenstein, Albert H., ass't ed. *Research Operations in Industry.* New York: Columbia University King's Crown Press, 1953. $8.50.

Hickman, W. Braddock. *The Volume of Corporate Bond Financing since 1900.* Princeton: Princeton University Press, 1953. $7.50.

Hilton, P. J. *An Introduction to Homotopy Theory.* Cambridge, England: Cambridge University Press, 1953. Paper 15 s.

Katona, George, and Mueller, Eva. *Consumer Attitudes and Demand, 1950–52.* Ann Arbor: University of Michigan, 1953. Paper, $1.50; cloth, $2.00.

Kenya Colony and Protectorate, Report on the Census of the Non-Native Population of, Taken on the Night of the 25th February 1948. Nairobi: Government Printer, 1953. Paper. 12/50.

Kuhn, H. W., and Tucker, A. W., eds. *Contributions to the Theory of Games, Vol. ii.* Princeton: Princeton University Press, 1953. Paper. $4.00.

Kyrk, Hazel. *The Family in the American Economy.* Chicago: University of Chicago Press, 1953. $6.00.

Longley-Cook, L. H., and Hooker, P. F. *Life and Other Contingencies, Vol. I.* New York: Cambridge University Press, 1953. $4.50.

MacNiece, E. H. *Industrial Specifications.* New York: John Wiley and Sons, 1953. $4.50.

Merriam, Ida C. *Social Security Financing.* Bureau Report 17, Social Security Administration. Washington, D. C.: U. S. Government Printing Office, 1952. Paper. $1.00.

Minnesota, University of, Industrial Relations Center. *Compensation Principles and Practices.* Research and Technical Report 13. Dubuque, Iowa: Wm. C. Brown Co., 1953.

Minnesota, Social Science Research Center of the Graduate School. *The Garrison State, its human problems.* Minneapolis: University of Minnesota, 1953. Paper.

National Bureau of Standards. *Hypergeometric and Legendre Functions With Applications to Integral Equations of Potential*

Theory. Applied Mathematics Series 19. Washington, D. C.: U. S. Government Printing Office, 1952. $3.25.

———. *Simultaneous Linear Equations and the Determination of Eigenvalues*. Applied Mathematics Series 29. (Edited by L. J. Paige and Olga Taussky.) Washington, D. C.: U. S. Government Printing Office, 1953. $1.50.

———. *Table of Natural Logarithms for Arguments Between Zero and Five to Sixteen Decimal Places*. Applied Mathematics Series 31 (reissue of Math. Table 10.) Washington, D. C.: U. S. Government Printing Office, 1953. $3.25.

———. *Tables of Bessel-Clifford Functions of Orders Zero and One*. Applied Mathematics Series 28. Washington, D. C.: U. S. Government Printing Office, 1953. Paper. 45 cents.

———. *Tables of Chebyshev Polynomials* $S_n(x)$ *and* $C_n(x)$. Washington, D. C.: U. S. Government Printing Office, 1952. $1.75.

Northcott, D. G. *Ideal Theory*. London: Cambridge University Press, 1953. Paper.

Organisation for European Economic Cooperation. *National Accounts Studies; Norway*. Paris: OEEC (Columbia University Press, official distribution agent in U.S.A.), June 1953. Paper. $2.00.

Quebec, Statistical Yearbook, 1951–52. Quebec: Department of Trade and Commerce, 1952.

Railroad Retirement Legislation, Report of the Joint Committee on. *Retirement Policies and the Railroad Retirement System. Part 1; Issues in Railroad Retirement: Part 2; Economic Problems of an Aging Population*. Washington, D. C.: U. S. Government Printing Office, 1953. Paper.

Schultz, Theodore W. *The Economic Organization of Agriculture*. New York: McGraw-Hill Book Co., 1953. $5.50.

Schweizerische Gesellschaft für Statistik und Volkswirtschaft. *Schweizerische Bibliographie für Statistik und Volkswirtschaft, 1950/52*. Zurich: 1953. Paper.

Tintner, Gerhard. *Mathematics and Statistics for Economists*. New York: Rinehart and Co., 1953. $6.50.

United Nations, Statistical Office of. *Commodity Indexes for the Standard International Trade Classification; Preliminary Issue*. Statistical Papers Series M, No. 10. New York: Columbia University Press official distribution agent in the U.S.A., April 1953. Paper. $5.00.

———. *Statistical Yearbook, 1952*. New York: 1952. Paper, $6.00; cloth, $7.50.

———. *Yearbook of International Trade Statistics, 1952*. New York: Columbia University Press official distribution agent in the U.S.A., 1953. Paper. $4.00.

United Nations, Technical Assistance Programme. *The National Income of the Philippines and its Distribution*. New York: Columbia University press official distribution agent in the U.S.A., 1952. Paper. 40 cents.

Virginia Polytechnic Institute, Bulletin of the, Vol. XLVI, No. 4. *The Department of Statistics and the Statistical Laboratory*. Blacksburg, Virginia: February 1953. Paper.

Walker, Helen M., and Lev, Joseph. *Statistical Inference*. New York: Henry Holt and Co., 1953. $6.25.

World Health Organisation. *Annual Epidemiological and Vital Statistics. Part I. Vital Statistics and Causes of Death, 1950*. Geneva: 1953. Paper. $5.00.

———. *Part II. Cases of and Deaths from Notifiable Diseases, 1947–49*. Geneva, 1953. $3.75.

Woytinsky, W. S., and Woytinsky, E. S. *World Population and Production; Trends and Outlook*. New York: The Twentieth Century Fund, 1953. $12.

Yokahama Mathematical Journal, Vol. 1, No. 1. Yokahama, Japan: May 1953. Paper.

Correction for publication listed in December, 1953 (vol. 48, no. 264):
Hood, William C. and Koopmans, Tjalling C., eds. *Studies in Econometric Method*. Cowles Commission for Research in Economics, 1953.

RANDOM DIGITS (12,876–15,125)

From *A Million Random Digits*, to be published by the Rand Corporation, Santa Monica, California.

Digits 8126–12,875 (erroneously labeled digits 9001–13,750) were published in Vol. 48, p. 931 (December 1953)

59748	72905	18532	10721	22029
06364	17893	86689	39755	39547
89772	55293	52849	31652	79655
24405	15409	33298	87632	61849
09657	36850	81569	83651	35795
48524	47778	80692	85476	23790
00740	36666	02680	21904	18370
11250	18463	86989	12625	30635
60923	68685	94994	87175	88318
40168	33501	59817	58830	00157
33842	89565	53359	13062	97614
97035	57905	09581	25343	17033
56429	73216	12342	14486	76624
63242	85960	56690	42093	35660
65814	21118	22140	36636	18291
02370	42160	73370	11944	85727
68037	41963	03874	44856	49762
00405	63369	55080	61880	26027
00234	14705	93418	94084	16525
91656	98079	52384	43306	31948
33557	87793	90857	10143	46726
91408	80220	05728	68890	46577
50106	50099	13722	18572	44024
57782	63951	53723	86853	63851
76162	71724	40028	94786	34457
95270	96584	81907	04055	53990
53445	67097	95523	66568	63632
77385	29911	65690	41178	47712
23854	34784	70950	54680	57811
72119	45668	03459	29870	78252
66397	85349	93335	86853	15860
71048	80847	75608	39646	90871
70355	81083	58581	44364	57468
53782	55926	64013	63562	41388
70088	12237	47838	46712	39848
81167	91259	14721	41014	97025
70284	82100	11669	02629	49845
21539	13042	76431	78515	02624
32397	74152	54102	80832	59979
35083	65927	25061	16625	77086

93225	08511	63096	37564	02738
73677	19193	14924	24365	83669
99744	13585	33746	86816	43028
09444	78700	02596	48353	26264
45911	39746	64722	75247	08432
75714	75249	95439	80714	73627
16479	53163	48071	28000	52002
08415	27236	52651	89059	45958
03415	57532	29981	61200	31962
89514	26883	69165	97237	12083
18232	92132	66537	70141	42854
19767	03409	11807	06565	66138
05000	29678	77726	73225	54753
87923	91889	68351	17712	75532
29209	90927	92929	92762	60413
49269	60819	77375	57886	47292
33891	00035	91300	14892	73653
23318	37253	38531	76489	49946
57517	67431	87124	57694	21660
41011	09285	61426	04658	54130
99670	73889	05601	32199	74923
26593	93877	45451	38735	42065
69108	59141	79502	69460	23108
64002	40998	50242	79738	96417
14710	20279	32747	49492	15399
98049	49717	70342	03953	83588
66769	69825	52614	24022	06278
74421	82990	27099	96177	41653
06664	19929	74427	49196	45007
24874	80825	92470	51514	52142
05910	12654	53630	13464	85697
24049	71670	43044	73649	23471
74493	20802	65843	92074	65712
32207	14097	66059	88577	99306
60961	57951	25637	93966	23073
95344	02606	74095	61133	04649
76052	82021	87676	91345	04967
46079	88514	08610	03685	15273
40706	66840	83003	51590	06230
24724	52989	71880	21952	09174
86643	75579	29276	10864	83179
56644	33354	31414	17643	92274
76278	53230	13953	76985	70959
76736	06558	64248	11907	29123
39263	54404	04355	64157	74407
01285	73719	80675	86178	56283
37007	02603	52673	44609	14843
06545	42195	44863	08871	89182
34989	95554	35562	85386	56252
00221	02446	64409	49406	99127

// JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 266 JUNE 1954 Volume 49

MEASUREMENT FOR ECONOMIC MODELS*

Stanley Lebergott
Bureau of the Budget

Three chief subjects are considered. (1) How accurate should data be for economic analysis via economic models, particularly input-output models? Preference is expressed for measurement of additional aspects of economic phenomena rather than broadside improvements in generally sound series. (2) How can models be developed to utilize imperfect data? Several suggestions are made: (a) examine the concepts and methods lying behind the basic data; (b) question models acutely sensitive to the inclusion of a single observation; (c) evaluate the economic meaning of models after empirical testing; (d) set up check models. (3) How can more adequate data be developed? A proposal is outlined for an integrated set of data on financial aspects of business, and for similar integration of data on employment and on consumer economic behavior.

More than half a century has passed since Marshall first sketched a system of equations which could be developed "until they embraced within themselves the whole of the demand side of the problem of distribution."[1] In that period the development of systems of equations, or models, has become a commonplace in economic study. Have our measurements kept pace with the demands which such systems place upon them? Three central questions can be set out for consideration:

1. How accurate should economic data be for model building?
2. How can models be developed to utilize imperfect data?
3. How can more adequate data be secured?

---

* A paper presented at the annual meeting of the Econometric Society, December, 1952. The opinions expressed are not necessarily those of the Bureau of the Budget.

[1] *Principles*, (London, Macmillan, 1916) Mathematical Appendix, Note xiv

I

How accurate should economic statistics be for model building? Clearly we do not wish them to be as accurate as possible: no one is prepared to devote the time and money required to achieve this end. We could, for example, immediately improve our estimates for gross national product, gross private domestic investment, new construction, personal consumption expenditures, incomes of unincorporated enterprise, personal savings, net realized farm income, index of prices paid by farmers, parity income index, and a host of other series if we had substantially better data on how much farmers spend for hammers and nails. We could reduce apparent inconsistencies between three sets of data on food consumption—Department of Agriculture estimates of food disappearance, the food component of the gross national product series, and the Census of Business data on food sales at retail—if we had a comprehensive study of food consumption outside private households. These projects, however, certainly lack the glamor and usefulness of many other projects; and it is most unlikely that funds could be secured to carry them out.

For general economic analysis, whose questions have been with us for a good many years, it is clearly difficult to demonstrate how accurate statistics should be. Nearly half a century ago, for example, Alexander Dana Noyes analyzed a period of recession, using statistics which reported a 50% decline in iron output, a 25% fall in textile output, a 55% rise in commercial failures and a 12% drop in railway receipts.[2] In 1886 Carroll Wright reviewed the course of the 1882–85 depression, noting the proportion of establishments idle and the number of business failures reported by R. G. Dun, commenting on steel and coal production series, on savings bank deposits and price trends, and concluding that the depression caused a "crippling of the consuming power of the people" but that "the volume of business transacted is not crippled comparatively to any such extent."[3]

Since these early days there have been obvious and welcome improvements in the accuracy of our data on industrial production and prices—to mention but a few. Yet if we review the reports in business magazines, the annual reviews of the Council of Economic Advisers and the Congressional Joint Committee on the Economic Report, can we state with assurance that such improvements have made any sig-

[2] A. D. Noyes, *Forty Years of American Finance* (New York, Putnam, 1909), p. 377–78.

[3] First Annual Report of the Commissioner of Labor, March 1886, *Industrial Depressions* (1886), p. 70.

nificant difference in the ability with which businessmen or government officials can evaluate economic trends?

Is the issue any clearer for model builders? Professor Morgenstern has pointed out as "obvious" that in large scale numerical operations such as those involved in inter-industry analysis "only very high quality data" should be used—given the extent of the numerical operations and their considerable cost.[4]

In actual practice how should such a prescription be applied? One of the dubious rows in the comprehensive and invaluable 190 by 190 inter-industry matrix which the Bureau of Labor Statistics recently prepared is that which distributed the output of "other repair services." The row for retail trade is only somewhat less dubious. Statisticians in the BLS and elsewhere have clearly recognized the inadequacies of these data and insisted on the desirability of improving them. Yet in the actual use of the model can we expect different policies to be pursued, or even suggested, if the aggregate demand for "other repair services" or for trade services differs by 1%—by 10%—by 20%? Probably not. On the other hand variations of 10% in the projections for copper demand might well be significant. Proceeding further, does this mean that all the coefficients indicating demands for copper by particular industries should be improved? Not necessarily. Of an estimated $1.3 billions in gross domestic output of copper in 1947, only some $200,000 was used by the industrial inorganic chemical industry. Clearly it is more important in terms of the final use of the matrix to improve not the coefficient for copper use by this industry but rather for use by the insulating wire industry, which consumed $345 millions—or even better, by the new construction industry which, on the crudest kind of data, was estimated to consume $162 millions.

Let us look at this from another viewpoint. As the size of the chart grows the size of the average coefficient diminishes—as does the likelihood that the error in estimating any given coefficient will significantly distort the estimates of output which are derived from models using the matrix. Hence the case for broadside improvements of data decreases in potency.[5]

It is particularly unfortunate that so many important studies of the

---

[4] Oskar Morgenstern, *On the Accuracy of Economic Observations* (Princeton, Princeton University Press, 1950), p. 39.

[5] Offsetting errors within a closed system will, of course, further minimize errors of estimate. However, there is no clear assurance that they will offset within the area of particular interest—e.g. demand for a particular industry's products—and *a fortiori* this argument does not apply in the usual open system.

stability of input-output coefficients, of the accuracy with which projections are made by models using input-output matrices, should have emphasized average errors or across-the-board accuracy.[6] Let us concede that it is useful to possess such overall measures. At the same time we must note that these measures are unrealistic—unrealistically harsh on input-output models as well as unrealistically easy on them. They are overly harsh because a complete improvement of all coefficients is certainly not to be achieved: it would cost as much as dredging a yacht basin or similar projects to which a distinguished senatorial economist has called attention.

Yet at the same time they are too generous insofar as they permit the implication that errors in estimating the total output of steel are no more forbidding than equal errors in estimating the total output, say, of trade. Clearly the difference is critical if such projections are not designed as mere mathematical exercises but intended for potential policy use.

Harold Barnett's valuable study makes it possible to empasize this distinction.[7] From his data we can contrast how the errors made in projecting the actual 1950 indexes of output for all industry groups compare with those for subtotals of separate industry groups. Let us set up two categories—one to include industries for which input-output projections would be of lesser interest since they would probably not be limiting elements in a full employment or mobilization situation—e.g. agriculture, food processing, etc.[8] The second group would include chemicals, metals transport, and the more limiting industries. In the consumption model the average errors in projecting 1950 indices by the 1939 input-output matrix were as follows:

All industries:	31.5 index points
"Less limiting" industries:	23 index points
"More limiting" industries:	36 index points

In the investment model the differences were even more marked; errors for the more critical industries were 41, or almost double the 21 for the less critical industries.

[6] Cf. *inter alia* Selma Arrow, *Comparisons of Input-Output and Alternative Projections, 1929–39* (The Rand Corporation, 1951).

W. W. Leontief, *The Structure of the American Economy, 1919–39* (New York, Oxford University Press 1951), pp. 216–18. These figures have been superseded by a comparison for 1950 outputs using the 1939 matrix in an unpublished memorandum by Professor Leontief and J. Fei. To some extent Waugh's analysis as applied by Christ is an exception. Cf. Part III of Carl F. Christ, "A Review of Input-Output Analysis," in *Conference on Business Cycles* (New York, National Bureau of Economic Research, 1951).

[7] Harold Barnett, *Specific Industry Output Projections* (The Rand Corporation, 1951). Table 4 gives deviation in index points from actual 1950 output levels.

[8] Agriculture, food processing, lumber, furniture, wood and paper printing, textile, apparel and leather.

The inference I would draw from all this with respect to the improvement of data for input-output and similar models is that we attempt increasingly to separate the sheep from the goats: industries with reasonably well behaved coefficients where erroneous projections of their output would involve small losses should be estimated as accurately as possible once—and thenceforth given second place in our attention relative to a very limited number of industries (such as steel foundries, primary copper, etc.) and the vast area of guessing the exogenous variables. To some fair extent this is what has been done in empirical research on input-output coefficients. We clearly do not require the mammoth sample survey of all manufacturing industries which has been proposed in the past in order to produce coefficients for all industries of "a very high quality."

All of which still does not solve our initial question. We may agree that neither improving the accuracy of all economic data nor achieving maximum accuracy for most data is a practical aim. But how accurate should any such data be? The question, I think, can only be solved *in ambulando*. There is no easy way of calculating the probable loss which results from errors in estimating economic data of the kind used by most econometricians. This is true because policy decisions do not usually rest on a single datum, a single series nor a single model. And it is even truer because errors in the choice of a model can have a far more critical impact than errors in the data.[9] It is up to the data producers, model builders, and data users generally to express their judgments on what levels of error are so intolerable that they believe more money should be spent for improving the data.

I should like to express a general preference for additional measurement series rather than improvements in existing series or data. Let us grant the obvious exceptions and assume, moreover, that the data are collected by persons with at least a minimum budget and minimum statistical competence. Then I think it can safely be said that knowledge about additional aspects of our economy will generally make a greater contribution to sound theory and sound policy than will improvements in the accuracy of existing series.

As one instance let me cite the patient and ingenious labors which produced Fabricant's index of manufacturing output.[10] The increases in accurate detail have been invaluable. Yet so far as concerns the totals for industrial production—and it is these which are used so

[9] The post-war forecasts erred because of the choice of model, the choice of consumption functions, because of ignorance about the way the labor market functions, business plans, business reaction to tax schedules and so on. Errors in the data were of small moment.

[10] Solomon Fabricant, *The Output of Manufacturing Industries, 1899–1937*. (New York, National Bureau of Economic Research (1940).

widely in economic models and policy decisions—the resultant index differs little from the one which Day and Thomas developed from far worse data with far less work.[11] Furthermore, neither index for total manufacturing would differ greatly in movement over the years from a very crude combination of output indices for textiles and iron and steel.

A similar consideration is apparent when one compares the recently revised and unrevised FRB production indexes, wholesale price indexes, etc.[12]

The substantial differences in series occur when a responsible investigator with a reasonable minimum of resources, cooperation and time first develops a series on a sound conceptual basis. Further improvements may be real; they certainly are hard won and usually expensive. They usually profit the model builder very little.[13]

In pressing for improved measurement in economic statistics the model builder would be well advised to give increased attention to the basic opportunity cost question. That question is this: do we improve our understanding of a particular economic phenomenon more by improving still further a particular series (or datum) or by adding new series on yet uncharted aspects of the phenomenon?

II

Given the present existence, and probable persistence, of errors in economic observation what is the model builder to do? How is he to build models for utilizing imperfect data? Models allowing for a random run of disturbances have been developed—well developed—in recent years and there is hardly need to recommend their use to econometricians. But if we are concerned with biased data, where acute errors are possible—and those not randomly disturbed—some alternative action is required.

A. The first requisite for the model builder is to examine the concepts and broad methods of estimation used in deriving the data which he incorporates into his model. I realize that this may be considered an undue burden; developing a sound model is a more than sufficient labor. Yet the alternative is even more unsatisfactory.

[11] E. E. Day and W. Thomas, *The Growth of Manufactures, 1899 to 1923* (Washington, Government Printing Office (1928), pp. 34, 94. With 1939 = 100, the 1899 estimates are 28 for Fabricant and (32) for Day-Thomas. Succeeding estimates are 34(39), 43(51), 51(55), 61(69), 54(54), 77(85), 82(88), and 87(88).

[12] The examples may be new, but the observation is an old one. Cf. Wesley Mitchell, *History of Prices During the War* (Washington, Government Printing Office, 1919), p. 28, where he indicates the similarity between the BLS index, based on 350 commodities, and the War Industries Board index based on 1,474 commodities.

[13] None of this, of course, bears on the question of what topics happen to evoke the abilities of research workers: if first class statisticians are willing to toil for years in the statistical salt mines, theirs is—and should be—the choice of mines.

Let us consider, for example, the ill charted and forbidding area of consumption projections. Klein is outspoken in asserting that "to many of us engaged in econometric work it became obvious in the second half of 1947 that the most serious deficiencies in the existing models lay in the consumption equation and in the group of relations serving to determine absolute prices."[14] Despite the wealth of data available on the subject, the relations between consumption and income continue to baffle projectors. On the one hand the budget surveys uniformly report that the proportion of income saved is greater at upper income levels than at lower ones. On the other hand Kuznets' data on capital formation seem to indicate that the aggregate consumption function has not changed perceptibly in the past sixty years when incomes have risen so.

In an acute discussion of the factors affecting consumption Professor Fellner has emphasized that

> The most obvious characteristic of the historical consumption function, as calculated from Professor Kuznet's estimates, is that it does not show a tendency to flatten out. It tends to linearity regardless of the varying population growth of the subsequent historical periods.[15]

A recent path breaking study on consumer behavior similarly notes that "the Kuznets' data do not show any trend in the savings ratio," suggesting that even if their level were incorrect they "allow us to make a judgment about the movement of the savings ratio."[16]

Reference to the basic sources indicates that to a very great extent this constancy exists because it was estimated that way. (And, incidentally, it was estimated that way because a sensible model indicated that to be the best method of estimate.)

More specifically: the consumption function relates consumption to net product. The latter consists of two segments: consumption and investment. Obviously consumption correlates perfectly with itself. But the investment segment also will reveal a high correlation with consumption because of the method of estimate. Let us consider in turn each of the components of the investment total.

a) *Inventory change.* Changes in manufacturing, trade and "all other" inventories—together accounting for more than half the total inventory change in most years—are estimated by constant ratios to output.[17]

---

[14] *Conference on Business Cycles* (New York, National Bureau of Economic Research 1951), p. 117.

[15] William J. Fellner, *Monetary Policies and Full Employment* (Berkeley, University of California Press, 1947) p. 56. Cf. also Paul Samuelson, in Seymour Harris ed. *Postwar Economic Problems* (New York, McGraw Hill, 1943), p. 33.

[16] James Duesenberry, *Income, Saving and the Theory of Consumer Behavior* (Cambridge, Harvard University Press, 1949), p. 56.

[17] Simon Kuznets, *National Product Since 1869* (New York, National Bureau of Economic Research, 1946), pp. 109, 110.

Now since a heavy proportion of total output is output of consumer goods, this method of estimate means that these inventories will necessarily fluctuate with consumer expenditures. Changes in inventories of livestock represent a further link, since livestock slaughter (a component of consumer expenditures) was estimated, in the original Department of Agriculture source, by applying ratios to inventories.[18]

b) Construction. The output of construction materials was used (with a weight of one, plus semi-durables with a weight of two) to interpolate estimates of consumer durable output between census dates.[19] This is a further link between the consumer and producer segment estimates.

c) Producer durables. The basic trends for producer, as for consumer durables, derive from the 1869, 1879, and other Census date totals estimated in William Shaw's comprehensive and fundamental study.[20] For most durable categories Shaw extrapolated 1914 and 1909 totals to earlier years by using a constant ratio to split Census group totals between consumer and producer durables.[21] The procedure assumes, for example, a constant ratio of consumer to producer goods in the output of furniture, sewing machines, foundry and machine shop products, etc. Thus 67.7% of all sewing machines were assumed to be household machines in every Census year from 1869 through 1909, with the balance classed as business use.[22] Other constant ratios (under 100%) were used for furniture, heating and cooking apparatus, appliances, cutlery, etc. For an important class of items Shaw made the reasonable assumption that none were bought by producers, all by consumers, at each of the Census dates. These include "family and pleasure" carriages and wagons, pianos, organs, rugs, glassware, books etc.[23]

So far as concerns the level of commodity flow results, or even many broad economic trends these decisions raise no question: a responsible estimator cannot arbitrarily vary ratios without some data to which to tie. But if we are concerned with something as precise as the constancy of the historical consumption function, we should decide how critical was the decision to assume that the rising number of apparel firms did not affect the proportion of all sewing machines bought by business, or the assumption that the growing number of restaurants and hotels did not change the proportion of all dishes and cutlery they bought,

[18] F. Strauss and L. Bean, *Gross Farm Income and Indices of Farm Production and Prices in the United States, 1869–1937*, (Washington, Government Printing Office, 1940), p. 106.

[19] *Ibid.*, p. 95.

[20] William H. Shaw, *Value of Commodity Output Since 1869* (New York, National Bureau of Economic Research, 1947)

[21] *Ibid.*, Table III.

[22] William N. Shaw, *op. cit.*, p. 161. For 1879–1909, 3.1% of the reported total is excluded as unfinished.

[23] *Ibid.*, Table III and Note B to Table III.

etc. When we come to use the data in economic models, we must not try to explain any constancy of savings ratios in final series if it rests ultimately on the constancy of arithmetic ratios used in the estimating process.

B. In addition to a careful examination of the methods of estimate a second step is to question models which are acutely sensitive to the addition or subtraction of a single observation—such as that for a single year or a single industry.[24]

A recent model which appears to be in this category is Modigliani's savings model.[25] Modigliani defines consumer saving as a function of the current year's income and highest previous year's income as measured by a cyclical income index. The model provides a beautiful fit for the 1921–40 period. If, however, we simply add the values for 1941, as provided in the appendix to Modigliani's study, we derive an estimated 1941 value so far in error that it cannot even be plotted on the chart he shows.[26]

His particular equation for allowing for the influence of previous years' incomes essentially states this—that for the entire decade of the 1930's the consumers continued to hark back to the halcyon 1929 level of incomes in deciding how much money to spend. However, the same model indicates that the consumer's frame of reference—say for deciding on his 1937 saving—would swiftly move forward six years from 1929 to 1936 if the 1936 total were a mere 4% greater.[27] (It would incidentally, take a courageous man to assert that Commerce 1936 income estimate is accurate to within 4%).

A fairly simple method to keep from implying that such drastic changes in consumer behavior occur when a single observation is changed slightly or added to previous ones would be to drop what is essentially a function based on ranking in favor of one based on a direct measure of magnitude. The latter would be less at the mercy of minor variations in source reporting or estimating techniques.[28]

Apparently, the basic equation for the demand for labor in Christ's

---

[24] Carl Christ wisely points out that if there were really a sharp change in important economic relationships then a model might properly be acutely sensitive to the addition of post-change data. One might add that economic relationships generally change so slowly that we should question both the data and the model, before we conclude that a sharp change in relationship has, in fact, occurred.

[25] Franco Modigliani, "Fluctuations in the Saving-Income Ratio: A Problem in Economic Forecasting" in *Studies in Income and Wealth*, (New York, National Bureau of Economic Research, 1949), Vol. XI.

[26] Franco Modigliani, *op. cit.* p. 380.

[27] The only consolation indicated is that since 1941 income levels were above any prior year 1929–40 the gain is defined as being a secular, and not a cyclical one. Consider the economic meaning of this. If per capita income rises $100, this is $100 of cyclic increase; if it rises $101, none of it is cyclical.

[28] A distributed lag approach was wisely suggested by Leontief; Modigliani noted that he had tried this and secured worse results. But less precise correlations based on due recognition of data shortcomings are to be preferred to the converse.

limited information version of the important Klein model III is almost equally delicate. By adding two observations to the 21 year run from 1921 through 1941—which had included short recessions, deep depression, and continued prosperity—Christ obtains a projected 1948 demand for labor which Klein reservedly calls "fantastic"—a comment echoed by Christ.[29] This, moreover, is no trivial or paltry incidental equation: it is fundamental to the model. It seems a striking fact that the disturbances which were trivial and random in Klein's limited information estimates become substantial and nonrandom in Christ's versions: they are negative for 1924, 1927, 1930, and other recession years, but positive (with one exception) for all other years.[30]

Christ assumes that the problem lies in the limited information method at least as much as with the equation. There is merit in this argument—but not enough. While the limited information solution actually blows up the standard error of disturbances even for the least squares solution is sufficiently disturbing—rising from .96 to 1.47 when Christ adds these few years to the original run. Klein favors errors in the basic data—particularly pointing to inadequacies in the BLS consumers price index. While the price data are dubious this is only part of the answer.[31] It would appear that the chief difficulty lies in the fact that Christ's calculations imply only a $9 billion rise in consumer expenditures from 1946 to 1947, whereas the Commerce data (published after his original work) report on $17 billion rise—reason enough to show the inconsistency between output and employment data which he and Klein discuss.[32] What is more important than the particular reason is the fact that any system which is so sensitive to the addition of several observations would seem to fall into the "handle with care" class.

C. There is a third consideration for model builders to bear in mind when using data—particularly imperfect data. And that is the need to evaluate the economic meaning of econometric models after empirical testing. Just as we will accept a sampling procedure which gives biased results provided the total errors are considered satisfactory so

---

[29] Carl Christ, "A Test of an Econometric Model for the United States, 1921–1947" in *Conference on Business Cycles* (New York, National Bureau of Economic Research, 1951), p. 124.

[30] *Ibid.*, p. 105.

[31] Using the BLS and BAE sources which Christ notes (*ibid.*, p. 91) one can estimate his price rise from 1946 to 1947 at 16%. This may be compared with the 10% gain shown by the deflator given the 1951 *National Income Supplement*, p. 146.

[32] Taking Christ's consumption data from his study (*ibid.*, p. 90) and applying the 16% rise computed above indicates a rise in consumption in current prices of $9 billions—as compared with a rise of $17 billions in the comparable Commerce consumption expenditure totals, (exclusive of imputed rent.) For the discussion by Klein and Christ, see *ibid.*, pp 98–101 and 114–15 especially. The error apparently arose in the use of regression procedures to extrapolate the Commerce figures.

we should prefer a model with greater resemblance to the economic world instead of one that has somewhat less error and distinctly less meaning.

It is instructive on this point to consider the recent study in which Tobin sets up a sensible model of the food market, combining budget data and time series data with equal parts of care, agility and aplomb.[33]

Tobin posits the following model of the retail food market:[34]

$$S_t = KY_t'^{a_1} Y_{t-1}'^{a_2} P_t^{\beta} Q_t^{\gamma}$$

where S^t is per capita food supply for domestic consumption, Y^1 is disposable income, P represents food prices and Q, all other prices.

Converting to a reduced form and rewriting, he developed his system as follows:

$$\log P_t = b_0 + b_1(\log S_t - \alpha \log Y_t') \\ + b_2(\log Y_t' - \log Y_{t-1}') + b_3 \log Q_t.$$

On a subsequent page, Tobin computes values for the parameters using various Agriculture and Commerce Department series for 1913–41. The values for each "b" are negative. The economic meaning of Tobin's model then seems to be this: When disposable income rises food prices tend to fall; and when food consumption rises more than past income-elasticity relationships would have predicted there will likewise be a tendency for food prices to decline.

This fairly odd pair of inferences suggests, I think, the desirability of evaluating the economic meaning of the model and its empirical parameters in some detail. The fact that the food consumption series is far less sensitive than the series for income or prices may offer an explanation. Its range is a mere 10% over the period Tobin uses it—whereas food prices and disposable income each had a 100% range.[35] This lack of sensitivity of the food consumption index suggests that some alternative measure of food consumption might properly be used —e.g. the deflated food expenditures figures of the Department of Commerce. A detailed comparison between the BAE and the Commerce indices (even after adjusting for such conceptual differences at the Commerce inclusion of services) strongly suggests that the two series do not move in consistent fashion.[36]

---

[33] James Tobin, "A Statistical Demand Function for Food in the U.S.A.," *Journal of the Royal Statistical Society*, Vol. CXIII, Part II, 1950.

[34] *Ibid.*, p. 130.

[35] This may well raise the question whether an identification problem is involved. Tobin carefully examines this question (*Ibid.*, pp. 135–6) and contends that it is not.

[36] Tobin infers that the failure to include in the series services in food distribution, would understate changes "in the supply of 'finished' foodstuffs." But nevertheless uses it—presumably as being reasonably sound. *Ibid.*, p. 131.

These are essentially negative cautions—though evaluating the methods by which his data were originally secured seems an elementary caution for any model builder to take, however pedestrian and however unrewarding such an inquiry proves to be. And analysis of the economic meaning of the resultant model—after the coefficients have been determined—seems equally urgent.

d) Positively, however, there is something which econometricians can do until that happy day when all economic data come from a single well designed survey producing data of only the highest reliability What can be done is to set up a check model—or models—to parallel the basic model. Such a model would incorporate related series or measurements in place of those used in their basic model. Clearly one would hardly seek to redo the 650 equations of the BLS 1947 matrix. Nor is this necessarily required. What we do need is a truncated model to give us some feeling for what changed inferences from the model would be produced by given errors in the original data. Let me be specific.

In recent models some fairly complex allowances have been made for changes in the distribution of income. To demonstrate that any of these allowances are real advances over, say, the simple procedure used by Tinbergen and Kalecki of distinguishing wage from non-wage income, we should have some measure of the adequacy of reported income distributions. One method would be to compare approximately similar distributions. Table I gives two distributions—one for 1934–35 and one for 1935–36.[37] The apparent differences between the 1934–35 and 1935–36 distributions are great—far greater, in fact, than the differences between surveys of slightly different populations separated by four years (1935–36 to 1939) of rising income.[38] A similar comparison for 1949, say, can be made between the Census and FRB surveys, and the differences are much smaller. In both instances the model builder might usefully use first one, then the other distribution to get an indication (*not* a precise measure) of the impact of survey variations on his conclusions.

A second method would be to make arbitrary adjustments in the

---

[37] Faith Williams and Alice Hanson, *Money Disbursements of Wage Earners and Clerical Workers in Five Cities in the West North Central-Mountain Region, 1934–36*, BLS Bulletin No. 641 (1939), p. 114. The concentration of the schedules in 1934–35 is noted on p. 341; A. D. H. Kaplan, *et al.*, *Family Income in Seven Urban Communities of the West-Central-Rocky Mountain Region*, 1935–36, BLS Bulletin 646, No. 1, p. 125.

[38] 1935–36 data from BLS Bulletin 646, No. 1, p. 110. 1939 data from 1940 Census, *Families, General Characteristics*, Table 59. These data relate to all families receiving wage or salary income. A distribution for all families without income other than wages & salaries would be virtually identical—for the $500 and over group discussed here.

data on the assumption of specified margins of error. For example, the income data reported in the BLS survey of 1941 aggregated about 10% less than control totals indicate was received.[39] More recent surveys have run from 5 to 10% low. What happens to our models if we allow for such understatements at each date, or if we presume the possibility that the underreporting of income decreased from the 1901 to the 1949 expenditure surveys?

TABLE I

DENVER WAGE EARNER AND CLERICAL WORKER FAMILIES WITH INCOMES OF $500 AND OVER PER CENT DISTRIBUTION BY INCOME LEVEL

	White nonrelief families*		All Families†	
	1934–35	1935–36	1935–36	1939
$ 500– 999‡	14	18	30	24
1,000–1,499	35	29	28	23
1,500–1,999	36	24	19	19
2,000–2,499	13	15	11	13
2,500–2,999	2	7	6	8
3,000 and over	0§	7	6	13
$ 500 and over	100	100	100	100

* Complete families, husband and wife native born.
† Families receiving $500 or more in wage or salary income.
‡ Families with incomes under $500 excluded from 1934 survey; also excluded from 1935–36 and 1939 data here for comparability.
§ Less than 1%

The measures need not be as close conceptually as this and cannot be in most instances. However, trends in the number of nonfarm employees (as measured by the Census Bureau's Current Population Survey) and in production (as measured by the FRB industrial production index) should be definably similar. Trends in manufacturers sales as reported by the Commerce sales series, and the FRB production index times the BLS price indexes should also bear a close relationship to one another and so on. The analyst can assess the result of using variant measures, deciding how much appears to be accounted for by conceptual differences, how much results from the fact that he chose one series rather than another.

[39] Selma F. Goldsmith, "Appraisal of Basic Data Available for Constructing Income Size Distributions," *Studies in Income and Wealth*, (New York, National Bureau of Economic Research, 1951), Vol. XIII, p. 285.

An additional reason for the use of check models is that we inevitably concentrate on one slice of reality, one aspect of the phenomenon being examined when we select any given series as an explanatory variable. It is frequently not clear whether that particular aspect is the most relevant, real or useful one to consider. Designating alternative measures helps triangulate the area of concern. As E. A. Goldenweiser remarked some years ago:

> no set of statistical series, to say nothing of any single series, is a sufficient basis for determining casual relationships on which economic policy can be predicted with safety. They are only indications of where one should look for the causes and interrelationships that determine economic events.[40]

The moral applies as well to the use of models for analysis.

III

How can we secure more adequate statistics both for policy determination and for economic analysis by economic models?[41] The simple answer would be to pass a law—not on statistics, of course, but on matters requiring statistics for their determination. Certainly, for example, the operations of the Securities and Exchange Act have developed better data on financial flows than even the combined talents of Crum, Epstein and others had been able to produce by the process of making bricks from straw. The passage of the Agricultural Adjustment Act brought support for agricultural income statistics of a kind that was rare even in the Department of Agriculture.

A more realistic answer to the question, however, must start from an answer to the question: what kind of data are needed for economic models? The answer, I believe, is essentially the same type of data as is required for general economic analysis except that it is much more urgent that the data be consistent. It is more of a problem for the model builder than for the general economist that the official series on plant and equipment expenditures is not consistent with that for sales in machinery and related industries; that one government series can report a 7% fall in auto inventories while another reports a 0.2% rise for the same period; that one series reports a drop in profits from

[40] "The Economist and the State," *American Economic Review*, March 1947, P. 5. Cf. also Charles D. Stewart and Loring Wood, "Employment Statistics in the Planning of a Full Employment Program," *Journal of the American Statistical Association* (September 1946).

[41] These purposes are not exclusive. In the preparation of the Federal Budget a set of economic projections are made for national income, prices, employment and other materials organized as a model. Similar materials are prepared by the staff of the Joint Committee on the Economic Report, Cf. Samuel M. Cohn, "Managing the Expenditure Side of the Federal Budget," a paper presented before the American Society for Public Administration, Washington Chapter, November 7, 1952. Joint Committee on the Economic Report, 82nd Congress, 1st Session, (1951) *The Economic and Political Hazards of an Inflationary Defense Economy*, Cf., especially Appendix C.

food manufacturing of 12% while another equally reputable series reports a rise of 20% that comprehensive reports on manufacturing employment for the same year may differ by several million depending on which series is selected; that employment data will be plant reports, production data may be company reports and profits data from consolidated company reports—each of these differences limiting our use of all three variables in model building, and so on through a catalog that is long, but not quite everchanging. To achieve such consistency, plus an improvement in accuracy, the following program areas deserve consideration.[42]

1. *Financial aspects of business.* Basic benchmark data can be provided annually for all of business, corporate and unincorporated from tabulations of the corporate, partnership and individual proprietorship income tax returns to the Bureau of Internal Revenue. Basic current data, used to extrapolate the benchmark data from quarter to quarter, can be secured from an expansion of the Financial Reporting Program now conducted jointly by the Federal Trade Commission and the Securities and Exchange Commission. That program, now restricted to manufacturing corporations, should be expanded to other industries, to unincorporated business. Now securing data on profit and loss, on balance sheet items, it could reasonably secure related data on business investment orders and plans for future sales and investment.[43]

The stakes for analysis here are first, accuracy; second, consistency; and third, a considerable increase in the number of observations. Instead of being restricted to one annual value for a year, only imperfectly reflecting a change in economic direction, the analyst could look forward to from 4 to 12 observations a year for a linked set of variables. The result should bring a much better knowledge of economic variation and relationship.

2. *Employment.* In the models developed by Hagen, the NPA, Fortune and other postwar projectors, in Klein's model III, and in the recent emergency model sponsored by the Air Force, production data, establishment reported employment data (from BLS), and household

[42] Richard Stone has argued persuasively for a single sample survey for business, one for consumers and one for government. Cf. his *The Role of Measurement in Economics* (Cambridge, University Press, 1951), pp. 57 ff.; R. Stone, J. E. G. Utting, and J. Durbin, "The Use of Sampling Methods in National Income Statistics and Social Accounting," in *Revue de L'Institut International de Statistique,* Vol. 18, No. 1–2 (1950); p. 31; and corresponding comments in reports of the UN Sub-Commission on Statistical Sampling. The consolidated surveys proposed here attempt to achieve the same goal of consistent data but at the same time make use of the mass of reliable data made available regularly in the U. S. from administrative reports.

[43] Administrative relationships are not particularly relevant here but it might be noted that in practice such a program would make use of such existing collection mechanisms as the Census Bureau's Current Business Reporting sample, etc.

reported employment data (to the Census Bureau) are all utilized. With labor a limiting resource in some projections, and unemployment an ominous resultant in others, the consistency of these data is particularly important.

Benchmark data for the basic BLS series come from the State Unemployment Insurance and Old Age and Survivors Insurance Systems. The reports to OASI, however, are simply tax reports to the Internal Revenue Service. It is worth considering whether the system of reports to the IRS might be so integrated that business would report annual payrolls and employment on their income tax returns.[44] If this were done we could then look to the monthly Bureau of Labor Statistics reports, like the FTC-SEC reports, as a means of keeping up to date on changes—with the additional consideration that the BLS data might be sought as in the past to apply to individual establishments rather than (as in BIR) to entire firms.

Such an integration of date on employment sales etc. would still exclude the Census Bureau's Current Population Survey. One approach would be to treat the list of establishments represented by workers who fall into the Current Population Survey as the sample for use by the Bureau of Labor Statistics.[45] By regular sample checks against establishment records of reports by individuals as to their employment status we have a method of reducing sampling differences between the BLS and Census surveys and removing characteristic errors in the responses secured by the population survey.[46]

3. *Consumers financial activity.* Moving from the sphere of business activity to that of consumers we are confronted with similar considerations. The analyst at present is likely to secure consumption functions from the well established BLS surveys. These functions may or may not be consistent with asset-savings relationships derived from the FRB Surveys of Consumer Finances, employment-income relationships estimated from the annual Census surveys or income elasticities for the consumption of food items as derived from data secured in

[44] BIR income tax returns do not now report employment, and permit the distribution of pay rolls under at least two deduction items.

[45] This proposal is discussed at greater length in the writer's "Labor Force Statistics: The Task Ahead," a paper presented at the 1950 annual meeting of the American Statistical Association.

[46] As has been emphasized by specialists in this field, errors in reports by the establishments would also have to be minimized—though this is presumably part of the regular program of the BLS. Administrative procedures would preclude sample changes more frequent than every year or so. However, if the survey is viewed as basically a means of interpolating between benchmark employment totals such an interval creates no real problem. The rate at which the sample should be changed must be judged in relation to the present rate of attrition and nonresponse in employment reporting by establishments. Accounting for the varying probabilities with which establishments would fall into the sample—i.e. what is the residence distribution of their employees; what is the correlation between employment characteristics of persons in the same family—is one of important technical issues which would have to be solved before such a proposition could be developed further.

Bureau of Human Nutrition and Home Economics and private surveys made for the Department of Agriculture. The survey data in turn are not fully consistent with the aggregates on income, saving and consumption that have been derived from business records and embodied in the national income accounts.

A program in this field, to be of value for economic models must provide consistent data, minimizing the number of measurements coming from completely different surveys. To achieve this purpose use can be made of the Census Bureau's Current Population Survey for securing an integrated body of data on incomes, expenditures, asset holdings, family employment status and demographic characteristics, as well as price and income expectations.[47] The gradual development of such a program would remove the considerable inconsistencies which now develop when the model builder incorporates measurements for many of these factors from the variety of surveys which now provide them.

A program in this field must provide reliable data, and if possible provide measurements which reflect variations in consumer behavior as variations in economic activity occur. In practice this would mean securing monthly reports from consumers on their economic activity in the previous month or week.[48] Such reports would facilitate accurate reporting by consumers, since the memory feats required would be minimized. No less important is the likelihood that their frequency would give us an improved knowledge of consumer activity. For example, interviews for the *1948 Survey of Consumer Finances* were conducted during January and February of 1948. Some 50 percent of the respondent interviewed in January expected a price rise. The grain market broke in February, and only 15 percent of those interviewed after the break expected prices to rise.[49] For the model builder the increasing number of observations at different levels of the nation's

---

[47] This proposal was discussed at greater length in a paper prepared for the 1949 annual meeting of the American Statistical Association, "The Validity of Interviews: Consumer Expenditure Surveys." To minimize respondent burden and nonresponse a system of replicated designs would be necessary: broad information would be taken from the main sample and subsidiary details from related samples.

Small scale studies by the Bureau of Human Nutrition and Home Economics suggest that some of the advantages of replicated samples may be delusive but that the procedure is practicable. Cf. Barbara B. Reagan and Evelyn Grossman, *Rural Levels of Living in Lee and Jones Counties, Mississippi, 1945, and a comparison of two methods of data collection* (October 1951) USDA Agriculture Information Bulletin 41, esp. Part 2.

[48] More than one period may be used, since reporting for long periods may provide reliable data for some classes of items (e.g., cars purchased), whereas briefer periods may be necessary for other items (cigarettes, milk, etc.).

[49] Federal Reserve Board of Governors, *1948 Survey of Consumer Finances*, Part I, Table 7. The relative change for farm operators and other groups as a whole was almost identical. Cf. James C. Davies, "Some Relations Between Events and Attitudes" *The American Political Science Review* (September 1952) p. 780

economic activity should increase the number of variables which he can include in his model, or test with any conclusiveness.[50]

Children's stories used to end with the magic words, "and they lived happily ever after." The prospect for econometricians, unhappily, is quite the reverse. Data will continue to be inadequate. What concepts lie behind the statistical measurements and what errors are hidden within them will continue to be issues for exploration. Econometricians will increasingly have to delay their more fascinating analytic work in order to ponder on the data and results, making certain that their findings do not merely quote errors, or assumptions inherent in the original measurements. They will in all probability find it essential to confirm each model with check models, making use of whatever scraps of evaluative information are offered them by data producers. In all this, however, there is at least one measure of consolation—the increasing frequency with which data producers feel obliged to check their data against control figures, to conduct post enumeration surveys, and to rely on scientific sampling so that at least measures of sampling errors will be available for data users. For both the producers of data and those who use them in economic models have a similar goal—to improve our understanding of economic change.

---

[50] It is, of course, possible to multiply observations even more cheaply, by a process of interpolation. This procedure produces useful data for some purposes. However, it does not really add to the number of independent observations of change through time. For a contrary view, in practice, Cf. Colin Clark's use of Barger's interpolated quarterly data in his "A System of Equations Explaining the United States Trade Cycle 1921 to 1941," *Econometrica* (April 1949).

TECHNICAL ASPECTS OF TRANSPORTATION FLOW DATA

R. TYNES SMITH, III

Bureau of Transport Economics and Statistics
*Interstate Commerce Commission**

THIS paper discusses the technical aspects of two problems relating to the 1 per cent sample of rail carload waybills currently being secured by the Interstate Commerce Commission. One concerns the selection of the sample itself while the other involves the use of the sample information for estimating a desired transportation statistic.

It is proposed to show that the present waybill sample is a simple yet powerful tool for transportation analysis. After a short historical review of the background of the problem in general, the sample selection procedure is described. This yields adequate but biased results so an additional adjustment is made which eliminates this bias and gives an efficient and representative sample. After this description an example of the straight forward manner in which the waybill data may be used to get answers to complex transportation problems is given by describing the technique used to develop a series of rate indexes. Finally several methods by which the standard deviations of these estimated indexes were actually determined will be outlined. But before proceeding a brief description of a waybill may be helpful for those readers who are not transportation experts.

A bill of lading is first prepared by the shipper which tells the carrier, among other things, the kind and amount of commodity to be shipped, and the points of origin and destination. The carrier prepares a waybill from this document which accompanies the shipment and which also provides the basis for assessing the transportation charges. After the shipment is delivered, the waybill is audited by the terminating carrier to correct the charges and is then filed. This audited document, therefore, provides a record of the actual transportation services performed and the charges made for that service. A copy is illustrated on page 238.

The use of waybill data in the analysis of transportation problems is not new. As a matter of fact, results of waybill studies have been used in proceedings before the Commission for at least fifty years. The first nationwide study was made in 1932, and the present waybill sample currently being secured by the Commission represents the first attempt to get continuing information in this manner. There has been a gradual evolution both in the type of data collected and in the technique of collection during this period.

* This paper has not been considered by the Commission.

The first individual studies were made to develop specific information and often covered only the movement of single commodities or only the traffic between particular points. There was no sampling problem because all of the bills for all of the traffic involved in a specific period of time were selected. The data were usually presented simply as observed facts for that period.

The first nationwide study was made by the Federal Coordinator of Transportation based on waybills covering one day's terminations in the year 1932. However, it was clearly recognized at that time that this sample could not be considered as a representative one so only limited generalizations were made from the relationships developed. The material was tabulated and no attempt was made to expand it as an estimate of the total traffic for a longer period than the actual day covered.

The first major attempt to obtain a countrywide representative sample for a whole year was made by the Board of Investigation and Research. Each Class I railroad was requested by that Board to supply it with certain information for all carload traffic terminated on its line on one designated day in each month of the year 1939. Different days in the month were assigned to the railroads in any one region so as to allow a better sampling of brief seasonal movements and the dates for parallel roads were staggered. Staggering of dates was arranged so far as possible to catch every week day during the week twice during year. This design was a major advance because the resulting sample could be considered as representative of the total traffic for the whole year.

Several waybill studies were conducted by defense agencies during the war. The War Department started a continuous sample of the bills of lading issued for its commercial traffic within the United States. This was a systematic sample which at the beginning was secured by simply counting the documents and making a selection in the proportion desired. Refinements were made in this procedure from time to time and a more efficient method was finally developed making use of the terminating digits of the bill of lading number.

It was found that a number of important characteristics such as type of commodity were related to the originating point and so stratification by origin prior to sampling greatly increased the efficiency of the sample. The bills of lading were prenumbered serially in large blocks and furnished in groups to the issuing office so that the desired stratification could be accomplished by arranging the bills according to their numbers. Analysis of the systematic selection of these bills led to the choice of a combination of the terminating digits as a selection device

which made the initial arrangement of the bills by number unnecessary but which preserved the advantages of origin stratification.

In 1946 the Commission started work on its waybill sample design and, of course, gave consideration to past experience in this field. Two practical methods of selection had been developed, the first based on all traffic terminating on certain days, and the second based upon terminating digits in the waybill number. Each of these plans offered peculiar advantages and disadvantages.

The selection of all traffic terminating on specified days was favored by many of the carriers because of its apparent simplicity. They also felt that there was some advantage in doing the necessary work involved in a short period of time in order to get it over with. However, there were some serious objections to this type of sampling, especially in connection with the ability to secure adequate representation of short seasonal movements and with the difficulty of determining the possible sampling error involved in estimates prepared from the sample.

The difficulty in determining standard errors of estimates based on solid days studies lies not in the fact that these are cluster samples, but that the clusters (days) are usually chosen on a judgment rather than a probability basis. Even if a suitable method of probability selection were applied the efficiency would be low because the seasonal factor of much of the traffic makes the between days variance of many important characteristics quite high. Since practical sample sizes lie between four and twelve days for a year, it is evident that estimates would necessarily be subject to relatively large standard errors. The author was recently told of a case where a sample consisting of four days, one in each quarter, was found to contain no observations of an important commodity known to be moving in large volume. A check revealed that there actually had been no movement on the selected days due to chance circumstances such as strikes and floods.

An alternative method of sample selection is based upon the terminating digits of the waybill number. This was a new idea to most of the carriers and many of them felt that it would present a more difficult problem than would the solid day's sample. It was also determined that there were a number of railroads using monthly numbering systems which would produce a bias in this type of selection if a standard combination of terminating digits were used for selection. Balancing these objections were the very important considerations of greatly increased efficiency in the sample, including the ability to reflect very small seasonal changes, and the possibility of making reasonably accurate estimates of the sampling errors of sample results. After careful con-

sideration of these and other factors it was decided to base the sample selection on the terminating digits of the waybill number.

A waybill sample could be selected either at the point of origin or at destination but the terminated bill is preferred because it contains complete information regarding the actual movement of the shipment and audited results of the charges assessed. However, this choice presents a problem when the selection is based on waybill numbers because these numbers are prepared at the origin points. There are several common methods of numbering waybills and the possible effect on the sample of these methods must be considered in the sample design.

All waybill series are numbered progressively to some point and then start over again at the initial number. The end of the series may be determined by some period of time, such as the end of a month, or by some number in the series itself such as the waybill number 10,000, 100,000, and so on. This latter type of series may be called a block system and it can easily be seen that the selection of all waybills with numbers ending in any given pair of digits from such a series will yield a systematic and unbiased one per cent sample of the waybills if the length of the block is some multiple of 100. Also, since it appears that the traffic characteristics of a shipment are independent of the terminating digits of the waybill number on which it moves, an unbiased sample of waybills based on these digits will also be an unbiased and representative sample of the shipments. It is immaterial in this case whether the selection is made at origin or destination.

The practice of starting waybill numbering series over again at the end of a period of time, such as each month or year, presents a different problem. It does then make a substantial difference whether the sample selection is made at origin or destination. An example of the effect at several small stations numbering their bills on a monthly basis will illustrate.

Suppose a sample of one out a hundred waybills is desired from a group of stations each issuing less than a hundred waybills per month. If the sample is to be selected at the origin station, there are a number of acceptable devices which could be used. A simple and direct approach would be to establish a separate register for checking off the bills as they were issued which would indicate each hundredth one for selection. This method, of course, is independent of the type of numbering system or of the volume of bills issued. Other methods can be devised using the waybill number but they immediately run into the difficulty that this number alone is not sufficient to make a one per cent selection. There will only be as many different numbers as there are

bills issued each month so that in the case of small stations some method of compounding must be used if a one per cent probability of selection is required. It is evident that this compounding device must be based on the volume of traffic at the issuing station and the waybill number initially used cannot be greater than the total issued in any one month.

These considerations, and the fact that the selection was to be made by the terminating carrier, dictated the choice of waybill numbers "1" or ending in the digits "01" as the best selection device. This permitted the issuance of standard, simple, and unambiguous instructions which are so necessary for successful operation. The compounding device required to secure a one per cent selection from the small stations was a procedure established at the Commission for subsampling the "1" bills from these small stations with a probability proportional to the average number of bills issued per month over a recent 12-month period. This method is roughly equivalent to weighting the received bills; the expected values will be the same although the variances will be somewhat higher because of the discarded observations. The difference is relatively small and is more than compensated for by increased simplicity in subsequent operations.

The excess bills over the required one per cent, which result from monthly and annual numbering systems, are an unnecessary cost to both the carriers and the Commission and so the carriers have been urged to change to the more desirable block system. Many have done so and consequently the problems associated with establishment of a proper sample design for small stations are becoming less important.

After developing and applying a correct sample design, it is evident that a satisfactory sample depends upon the complete and accurate selection of the designated waybills. These waybills are selected by the reporting carriers in over one hundred offices through the occasional efforts of perhaps a thousand individuals. This factor introduces possibilities of error which require constant policing to assure the desired accuracy. Fortunately, the Commission receives another report on carload traffic which can be compared to the waybill sample as an effective check for completeness and accuracy.

This report is known as the quarterly freight commodity statistics and shows for each railroad the total number of cars terminated by that road for each of the 261 carload commodity classes divided between traffic which originated on the reporting road and traffic which was originated on other carriers and was delivered to the reporting road for termination. Each quarter the waybill sample is tabulated in

this same manner for each reporting carrier and the results, commodity by commodity, are compared to 1 per cent of that road's freight commodity statistics. Any significant discrepancies noted in this comparison are carefully reviewed and the causes for such discrepancies are determined. It has been interesting to note that some of the discrepancies have been due to errors in the commodity statistics rather than in the sample and so in these cases the sample has actually been used as a quality control for the 100 per cent report. An idea of the effectiveness of the control of the sample may be gained from the fact that in each quarter there are in the neighborhood of 20,000 individual comparisons. The sample itself includes about 75,000 observations.

The selection of waybills on the basis of the waybill number provides in effect for a sample of every hundredth bill issued by each station. This is actually true in the case of stations issuing bills on a block system and in the case of stations issuing many hundred bills per month and is approximated at the others. Consequently the sample is quite efficient because of the relationship between the originating station and important transportation factors such as the commodity and territory. The periodic comparison of the sample to control totals helps reduce selection errors and assures its validity. The sample has found many uses in varied fields and its application to one transportation problem will be discussed, but before doing so it might be well to consider the foregoing procedures in the light of modern sampling theory [1].

The universe considered here is the totality of rail carload shipments terminated by Class I railroads in the United States. The sampling unit is the waybill, which in general covers a single carload, and the frame is the complete series of numbered waybills. These numbers are assigned at the origin station so that when a block numbering system is used selection based on the terminating digits "01" will provide a systematic sample of every hundredth waybill issued, stratified by origin. Since the shipment itself is independent of the terminating digits of the waybill number this will result in an unbiased, random—and therefore representative—one per cent sample of carload shipments.

A modification of this design is necessary in the case of small stations numbering their bills on a monthly system. Where at least one but less than 100 bills are issued per month the selection of the "1" bills provide a sample consisting of the first shipment in the month for each station. This, therefore, is not necessarily a random selection except in the sense that there is some chance element in whether or not a particular ship-

ment will be the first one. The essential random element is provided by a further process of subsampling these initially selected bills.

The "1" bills from each of the small stations over several months form a series which approximates a systematic sample with a sampling interval equal to the average number of bills issued per month. The desired interval is 100 and this in turn can be approximated by discarding a sufficient number of bills so that the resulting average interval is 100. The bills to be discarded are selected by a chance process with appropriate probabilities so the final sample is a random one. This permits valid estimates of standard errors, an example of which is given later.

There is a problem of non-response due usually to occasional failure to select all the required "1" and "01" bills. These failures are essentially random in character and do not appreciably affect computed averages or ratios although of course they do introduce a negative bias in estimates of aggregates. The quarterly comparison to the 100 per cent count of the commodity statistics report provides an excellent control and good progress is being made in the establishment of accurate mechanical checks by the reporting carriers which reduce the non-response error.

Many diverse uses have been made of the Commission's waybill sample but it is proposed to examine the technical aspects of its use in only one problem. This problems, however, is one of sufficient complexity to indicate the value of the sample as a research tool.

The need for an index which would measure changes in rail carload freight revenue resulting solely from increases or decreases in freight rates has been recognized for years. Experimental indexes for various commodities were developed by the Commission but the methods used were too costly to permit any extensive expansion of these series. The Commission's 1 per cent waybill sample provided a new approach to the problem and offered the hope that annual indexes of many freight rates could be prepared and kept current with a relatively small expenditure of time and funds. The method finally adopted uses, with minor modifications, information contained in the regular releases of waybill statistics.

The regular Commission waybill statistics are tabulations of punch cards prepared from the waybill samples reported by the terminating carriers. The illustration on page 238 shows facsimiles of a typical waybill report and the detail and waybill cards prepared from that report. It also gives a brief description of the processing of the bills and cards.

Mileage block statistics have been developed by classifying the

traffic according to traffic categories determined by the commodity class, short-line length of haul, type of rate, and territorial movement. These are the statistics used in the preparation of the freight rate indexes. It should be noted that the rates, or prices charged for transportion, are related to each of these factors and they were chosen with a view towards making the items within each resulting traffic category as homogeneous as possible. The sample shipments for each year fall into about 30,000 of these traffic categories.

The indexes are constructed by comparing the tonnage and average revenue per ton of matching categories in the base and comparing year. Two sets of revenue figures are prepared from these data. The first consists of the actual revenue for the base year which, of course, is simply the total of the revenues for each of the comparing categories in this year. A second set of revenue figures are obtained by applying the average rate per ton for each category in the comparing year to the tonnage which moved in the same category for the base year. This produces a figure which represents the revenue which the traffic moving in the base year would have produced at the average rates in effect in the comparing year. The ratio of these two revenues yields the desired index figure.

It is evident that factors other than changes in freight rates might affect the average revenue per ton for a given traffic category. For example, the average length of haul within the mileage block assigned to the traffic category might be greater or less in the comparing year than in the base year. Similarly, the specific commodities which moved within the given commodity class might be different and take different rates. Therefore, the average revenue effect of such differences must be small as compared to the effect of rate changes if the resulting indexes are to be a satisfactory measure of changes in rates. Careful study of this problem does indicate that this requirement is substantially satisfield. As noted before, the traffic categories were chosen so as to be as nearly homogeneous with respect to rate characteristics as practicable. The milage blocks are shorter where changes in rate progression are most rapid. The various territorial movements are kept separate as are movements on interstate and intrastate rates and there is a further classification by commodity class and type of rate. Consequently, the area of fluctuation in average revenue from causes other than changes in the rate is relatively small in each of the traffic categories.

It is reasonable to expect that even when such fluctuations as these do occur, they are as likely to affect the results in one direction as another so that there is a tendency for the errors to compensate rather than to accumulate. In consequence, the net effect is that the changes in

average revenues noted by this method will usually reflect quite closely the revenue effects of actual changes in rates. Therefore, the indexes will be a measure of the average changes in rates. An illustrative page from the published statement is shown in Table 1.

These indexes are based upon only one of the 100 similar 1 per cent samples which could have been selected from the waybills representing the total carload traffic. Presumably indexes computed in exactly the same manner from each of the 100 different samples would have produced 100 slightly different results. Each of these in turn would have been an estimate of the index which could theoretically have been produced from a 100 per cent sample consisting of all of the waybills. This condition, of course, immediately raises the question of how much does a particular index computed from the sample which actually was selected differ from the theoretical index which could have been prepared from the total of all waybills.

There are a number of methods by which an estimate of the standard deviation of any computed index could be obtained but in particular there are two relatively simple procedures which were actually applied to the computed rate indexes and which have been used for similar determinations of the confidence which could be placed in other estimates prepared from the waybill sample.

It was noted that you would expect somewhat different results from computations made on each of the 100 possible 1 per cent samples which could have been drawn from the total of all waybills. If the variation in these estimates could be known, it would be easy to calculate the standard deviation for one of them. Obviously, it is not possible to get such figures as these but an extension of the reasoning can be made to the sample which is available. If this sample is divided up into, say, ten parts in a manner similar to the method used in the selection of the original sample, then the variation observed in separate estimates made from each of these ten parts can be used to estimate the possible variation in the estimate prepared from the whole sample [1]. If $X_1 \cdots X_{10}$ are the ten subsample estimates then the variance of the total sample estimate $\overline{X}$ is approximately:

$$S_{\bar{x}}^2 = \frac{1}{10}\,\frac{1}{9}\sum_{1}^{10}(X_i - \overline{X})^2$$

or, using the range $R = X_{max} - X_{min}$ the standard deviation can be estimated from the relationship [2]

$$S_{\bar{x}} = \frac{.325}{\sqrt{10}}R = .103R.$$

TABLE 1

INDEXES OF AVERAGE FREIGHT RATES FOR COMMODITY GROUPS AND SELECTED COMMODITY CLASSES*
(1950=100)

Item	Index 1947	1948	1949	1950	1951	1952	Per cent Increase 1952 Over 1947	Approx. Standard Deviation
All Commodities	80	93	99	100	102	109	36	.5
Group I—Products of Agriculture	80	93	98	100	102	108	35	.5
Class 001 Wheat	80	93	95	100	102	108	35	1.0
003 Corn	75	94	95	100	102	109	45	1.0
033 Cotton in Bales	82	92	98	100	102	112	37	1.0
† Oil Bearing Crops	78	91	99	100	103	111	42	4.0
† Fresh Fruits	83	95	99	100	101	106	28	.5
† Fresh Vegetables	82	95	99	100	101	106	29	.5
085 Potatoes, Other Than Sweet	79	94	99	100	101	107	35	1.0
101 Sugar Beets	86	96	103	100	106	110	28	3.0
199 Products of Agriculture, N.O.S.	88	94	105	100	101	116	32	3.0
Group II—Animals and Products	77	93	99	100	102	110	43	.5
Class 203 Cattle and Calves, S.D.	79	94	98	100	103	112	42	1.0
215 Meats, fresh, N.O.S.	73	90	98	100	103	110	51	1.0
Group III—Products of Mines	83	91	98	100	102	108	30	.5
Class 301 Anthracite Coal, N.O.S.	81	90	99	100	101	106	31	.5
305 Bituminous Coal	81	89	98	100	102	107	32	.5
307 Coke	83	91	97	100	103	108	30	1.5
309 Iron Ore	88	95	99	100	103	110	25	.5
323 Clay and Bentonite	84	95	98	100	104	111	32	1.0
325 Sand, Industrial	77	91	98	100	105	113	47	1.0
327 Gravel and Sand, N.O.S.	90	98	102	100	102	107	19	1.0
329 Stone and Rock: Broken, Ground and Crushed	85	93	100	100	103	108	27	1.0
331 Fluxing Stone & Raw Dolomite	75	88	98	100	104	109	45	1.0
337 Petroleum, Crude	81	96	99	100	102	109	35	1.5
339 Asphalt	77	94	100	100	101	110	43	1.0
341 Salt	86	93	98	100	102	108	26	1.0
343 Phosphate Rock	81	95	100	100	104	110	36	1.5
399 Products of Mines, N.O.S.	82	98	100	100	103	112	37	2.5
Group IV—Products of Forests	79	93	98	100	102	109	38	.5
Class 401 Logs, Butts, and Bolts	83	93	99	100	103	108	30	1.5
403 Posts, Poles, & Piling Wooden	79	92	97	100	101	111	41	1.5
409 Pulpwood	88	98	100	100	104	110	25	.5
422 Lumber, Shingles, and Lath	78	92	98	100	102	109	49	.5
499 Products of Forests, N.O.S.	83	95	100	100	104	112	35	2.5

* Only classes having more than approximately 1,000 cars in the sample are shown separately. The group indexes cover all classes in the group.

† These groupings of individual commodity classes are shown at the request of the Department of Agriculture. "Oil bearing crops" include commodity classes 037 to 047, 097 and 105; "Fresh fruits", classes 049 through 069; "Fresh vegetables" classes 077 through 089 (including 085 which is also shown separately).

This latter estimate is only about 85 per cent as efficient as the former but has the great advantage of simplicity and ease of computation.

The first estimates of the standard deviation of the rate indexes were made by following this procedure in principle. It was recognized, however, that the preparation of ten separate sets of indexes from ten subsamples would be too great a task for the limited staff available. Consequently, the individual waybills in the sample were divided into only five groups. This division was made as nearly as possible on the same basis as that by which the initial sample was selected. Each of these five groups was classified according to traffic categories and five different sets of indexes were computed. The variation in these indexes then yielded estimates of the standard deviation for the indexes prepared from the total sample. This method, however, proved to be an even greater task than initially contemplated and a less accurate but substantially less costly procedure was developed.

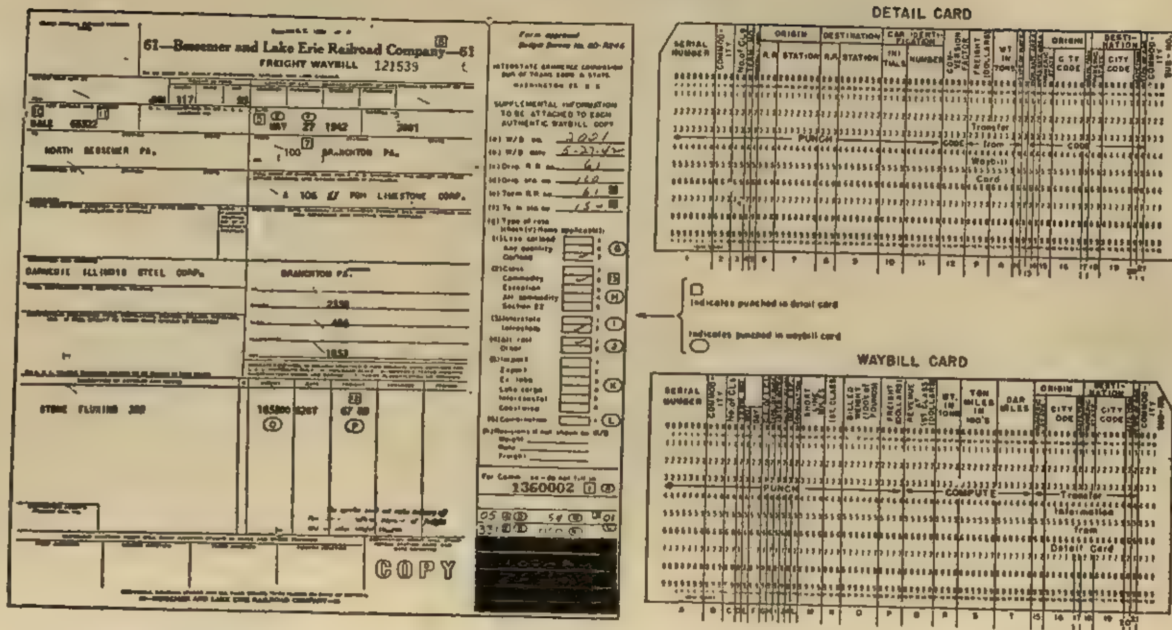
The second procedure made use of the same mileage block cards that were used in developing the initial indexes. These cards were randomly distributed into ten groups and the indexes prepared as before for each of the groups. The variation in these indexes then yielded an estimate of the standard deviation of the index prepared from the total of the mileage block cards. While this procedure is not as accurate as the one using the individual waybill cards, the error is on the safe side because there is a bias which tends to indicate a larger value of the standard deviation. However, the substantial savings resulting from the use of mileage block cards instead of the individual waybill cards more than offsets the loss in accuracy.

Incidentally, the mile block cards were randomly distributed into ten subgroups by means of a one column sort on the units position of the revenue field. The digits in this field are equivalent to the sum modulo ten of the units digits in the waybill cards included in each block. It has been shown [3] that such a process of "compound randomization" will yield sequences approaching equal probabilities for each digit, even when the generating sequences are quite biased. Consequently it is permissible to use aggregate fields of summary cards in this manner even when the same fields in the detail cards would be unsatisfactory. Extensive tests have been made on a table of random digits [4] produced in a similar manner from waybill data which indicate that considerable confidence can be placed in the effective randomness of the results.

The results of the two procedures for estimating the index standard deviations were in substantial agreement although individual examples of wide variation did appear. These were usually due to unequal distributions of traffic in the mileage block cards which of course is one

INTERSTATE COMMERCE COMMISSION
BUREAU OF TRANSPORT ECONOMICS AND STATISTICS

CONTINUOUS CARLOAD WAYBILL STUDY
CARD FORMS AND PROCEDURE



Immediately upon receipt the waybills are reviewed for completeness and legibility and the number "1" waybills issued by stations having monthly numbering systems are subsampled to eliminate the bias resulting from the numbering system. The short-line distance and first-class rate between origin and destination stations are then checked and the bill is coded for commodity, month of termination, and number of carloads. It is then serially numbered and two cards are punched as shown above.

The detail card is used for mechanically coding the origin and destination and type of car. It is first matched against a station master file which contains a city, state, and geographical area code which is reproduced into the detail card for both origin and destination. The geographical area codes are used to develop the territorial movement which takes into account the direction of movement. The detail card is then matched against the car initial and number master file which contains the type of car code and which completes the coding.

The waybill card is the basis for computing revenue at first-class, weight in tons, ton-miles and car-miles. When these multiplications have been completed and verified the two cards are matched against each other and information transferred from one to the other as indicated.

Tabulations can be prepared from either or both cards. Through the use of summary cards, additional computations and analyses can be made for special characteristics of traffic.

of the prices which must be paid for using this approximation method. However, as was expected, the discrepancies were usually on the side of overestimating the error and so the less costly procedure is considered a safe one for determining the confidence which can be placed in the estimates.

The preceding discussion has covered the selection and adjustment of the carload waybill sample, and the method used to check its completeness and accuracy by comparison with the carriers' freight commodity statistics report. It has shown how the regularly published waybill statistics can be used, with minor modifications, to get the answer to a complex transportation problem in a simple and straight forward manner. As a final example several methods for estimating the standard deviations of sample results were developed. In each case the techniques used were simple and direct. This is characteristic of the current waybill sample and illustrates its value as a tool for transportation analysis.

REFERENCES

[1] Deming, William Edwards, *Some Theory of Sampling*, New York, John Wiley & Sons, Inc., (1950).

[2] Dixon, Wilfrid J., and Massey, Frank J., Jr., *Introduction to Statistical Analysis*, New York, McGraw-Hill Book Co., Inc., (1951).

[3] Horton, H. B., and Smith, R. T., "A direct method for producing random digits," *Annals of Mathematical Statistics*, 20 (1949), 82–90.

[4] Horton, H. B., and Smith, R. T., *Table of 105,000 Random Decimal Digits*, Washington, D. C., Interstate Commerce Commission, (1949).

RESPONSE ERRORS IN THE COLLECTION OF WAGE STATISTICS BY MAIL QUESTIONNAIRE

SAMUEL E. COHEN AND BENJAMIN LIPSTEIN
U. S. Bureau of Labor Statistics

THE literature on response errors in mail questionnaires has for the most part been devoted to the personal characteristics of the individual, e.g. age, education, attitude toward social questions, etc. Little attention appears to have been given to response errors where the data are collected from firms or organizations. It is doubtful, however, that a sharp separation can be made between response errors of personal characteristics and data on business operations. In the small firm, elements of prestige may exist with respect to exaggerating sales, production, or employment. The large firm on the other hand may be reluctant to disclose the extent of its operations for competitive reasons, or in government surveys may fear that the information may be used for other than statistical purposes. Further, during periods of economic control, some business firms may consider it advantageous to bias reports in their own favor for future administrative action. Thus, it appears that a parallel may exist between response biases of individuals and the response bias of firms.

The problem errors of response of firms to a mail questionnaire was recently studied in a wage survey of the rainwear industry conducted by the U. S. Bureau of Labor Statistics. Two kinds of response errors were examined, 1) coverage errors and 2) errors in the reporting of wage data. Coverage error as part of response error exists in this instance because the firm classifies itself as being in or outside of the industry. It includes firms that were incorrectly classified as being in the industry as well as those which were incorrectly excluded.

Errors in reporting of wage data include such errors as incorrect reporting of wage rates, inclusion of workers outside the scope of survey (e.g.—office workers), and erroneous division into the auxiliary[1] or non-auxiliary class.

The survey was initially made during July 1951 for a June 1951 payroll period, to provide data for aiding in the determination of the pre-

[1] For the purpose of this survey auxiliary workers are defined as those workers engaged in one or more of following occupations: position marking, shade and size numbering, bundle tying, bundle ticketing, matching and pairing, basting, pulling, hand trimming, cleaning, turning; floor boys and girls, porters, examiner's helpers. Non-auxiliary workers consist primarily of cutters and sewing machine operators.

vailing minimum wage in that industry.[3] Responses were obtained predominantly by mail questionnaire. A subsample of the non-respondents to the mail questionnaire was contacted in person or by telephone so that an unbiased estimate of the wage distribution of the industry could be prepared. Subsequently, questions were raised as to whether mail collection was appropriate, and whether the data presented were acceptable for use in minimum wage determination. Accordingly, a field resurvey was made to check the survey results and to determine whether mail questionnaires are feasible in surveys of such industries as rainwear, where piece-work payments prevail.

The original respondents were given no advance warning of the resurvey. Difficulties generally encountered in studies of response error because of uncertainty in identifying the reporting unit were not present in this survey. It was possible to identify the reporting units as being identical in both surveys—something not always possible with families or households.

The variables measured in both the survey and resurvey were subject to objective determination from payroll records. Thus variations in reporting would be attributed to indifference, misinterpretation of instructions, arithmetic errors or purposeful concealment. These items cannot be isolated in most studies of survey techniques and reporting practices. For example, differences in reporting educational attainment can be attributed in part to the uncertainty of the definition. A very common source of variation in reporting, age, is generally not subject to any definite verification. All that is available as a rule are two replies by the same respondent, both of which may be in error.

SURVEY METHODS

The universe of establishments in the rainwear industry was defined to include all establishments whose value of product of rainwear was 50 per cent or more of their total production in 1950 or who maintain separate rainwear departments. An industry of this type is not easily defined or isolated from the remainder of the needle trade industries. A comparatively small number of firms are continuously in the industry from year to year. For the most part, firms shift in and out of this industry with the fluctuations in product demand. The plant

[3] The Walsh-Healey Act specifies that employers engaged in government contracts exceeding $10,000 must pay the prevailing minimum rate of pay for such work. The Secretary of Labor is assigned the responsibility of determining the prevailing rate of pay in question. Administratively, this work is carried out by the Wage and Hour and Public Contracts Division of the Department of Labor. The Bureau of Labor Statistics is frequently called upon to conduct surveys of wage rates in the industry. The data collected in these surveys are used as an aid in this wage determination.

equipment of firms producing rainwear is not specialized, but rather is equally usable for many other needle trade products.

The initial listing of firms comprising the industry was obtained from Unemployment Compensation listings, supplemented by lists of firms provided by the International Ladies' Garment Workers' Union and the Amalgamated Clothing Workers Union.[3] The first wave of questionnaires was mailed to all listed firms (633). A second wave of schedules was sent to all nonrespondents. The 353 responses to these two mailings consisted of 84 firms being defined as within scope (firms within the rainwear industry) and 269 firms as being out of scope. Of the nonrespondent group, a sample of 70 firms was selected for direct contact by telephone or personal visit. The non-response sample produced 19 firms within the scope of the survey and 51 out of scope.

TABLE 1

SAMPLING PROCEDURE OF ORIGINAL SURVEY

Stages in the Sampling Procedure	Number of Schedules	In Scope	Out of Scope
Total mailed	633		
Responded	353	84*	269
Did not respond	280		
Direct contact	70	19†	51
Total received	423	103	320

* 66 schedules usable—rejected schedules failed to provide wage distributions or other essential data.

† All schedules usable.

The resurvey consisted of two parts, 1) a sample check of firms reported initially as out of scope, and 2) the collection by personal visit of data initially obtained by mail.

Part (1) of this investigation consisted of a sample of 42 firms in New York City classified as out of scope on the basis of mail returns in the initial survey. These 42 firms were rechecked by personal visit. Of 17 firms initially reported as out of business, of the sample of 42, it was found that 10 were completely out of business, one was manufacturing another product, 4 were defunct rainwear subsidiaries of firms still in other lines, and 2 were in the rainwear business elsewhere and under different names (included in the universe under the new name).

[3] Past experience has indicated that lists compiled in this manner contain most of the firms in the industry. An exhaustive investigation of the completeness of the list would be financially prohibitive, as it would involve a detailed investigation of a number of related industries.

Of the 25 remaining firms in the out of scope sample—only 3 could be classified properly in the rainwear industry as defined.

On the basis of this sample survey, it appears that the Bureau underestimated the number of rainwear establishments in New York City by about 15 or approximately 20 per cent.

Most of these 42 establishments at some previous time manufactured rainwear. However, a general decline in the volume of such production had resulted in rainwear no longer being the principal product or the discontinuance of a separate rainwear department. Such shifts had been taking place since 1946 and explain in part the divergence of the survey data with respect to number of establishments (1951) from that of the 1947 Census of Manufactures.

TABLE 2

RESULTS OF RESURVEY OF FIRMS CLASSIFIED AS OUT OF SCOPE IN ORIGINAL SURVEY, NEW YORK CITY

Classification	Number of Firms	
	Original Survey	Resurvey
Total Sample of Firms	42	42
Out of Business	17	10
Change of product		1
Discontinued rainwear dept.		4
Operating, under different names:		
Included in universe		2
Manufacturing other products	25	22
Manufacturing rainwear		3
Total out of scope	42	39

Despite the under-representation of the Middle Atlantic region (as suggested by the New York City results) in the Bureau of Labor Statistics survey, no great difference in the Nation-wide statistics could be generated by such a deficiency. Under the extreme assumption—that the bias of omission occurs in that area alone—the increase of the Middle Atlantic employment raises the Nation-wide median by 1 cent and decreases the percentages of workers in the lowest class intervals (75 cents–80 cents) by no more than three-tenths of 1 per cent. Actually, an understatement of 20 per cent in number of firms in New York City

and in turn in the Middle Atlantic Region means less than a 20 per cent understatement in employment since firms in this region are smaller than those in other regions. Such field work as was done in New York is very expensive to do elsewhere because of the scatter of the remaining establishments. A small sample in other areas would be practically useless because of the apparently low proportion of bona fide rainwear firms erroneously excluded.

Do errors of the universe, such as we have just described, form an effective argument against making mail surveys in such industries as rainwear, where the establishment turn-over is so great? Without the expenditure of large sums, it is likely that the error in any field survey would be even greater. In order to obtain 85 usable schedules, it was necessary to make an initial mailing to more than 600 establishments, most of which did not fall within the scope of the survey. The field experience would not be appreciably better than this, judging from the canvass made in New York. Such a canvass by personal visit would be considerably more expensive. Hence the alternatives presented are (1) a mail survey adjusted for non-response, or (2) no survey at all. The extent to which correction should be made depends on the use to be made of the survey.

In the second part of the resurvey, an attempt was made to visit all firms not originally interviewed in person. However, visits were not made by the Bureau to 13 establishments which were included in a field survey made by Dr. Lazare Teper of the International Ladies' Garment Workers' Union. Dr. Teper provided the Bureau with data collected in his survey.

Wage data collected in the resurvey were obtained directly from payroll records. Schedules were obtained from 53 of the 57 firms visited by by Bureau of Labor Statistics. One firm employing about 100 employees refused to cooperate in the resurvey. The other three, employing a total of 28 workers, were either out of business at the time of the resurvey, or their records were unavailable.

The retabulations are thus based on 15 schedules obtained originally by personal visit, 13 obtained by the ILGWU, and 53 obtained in the resurvey.[4]

All schedules were given the same weight in the retabulation as in the original survey. The final tabulations are therefore the equivalent of a complete field survey.

[4] Four of the original 19 schedules obtained by direct contact as mentioned in Table 1 were obtained by telephone. These were included in the 53 firms resurveyed.

COMPARISON OF RESULTS

In comparing the results of the two surveys, it is assumed throughout that the data collected by personal interview are without errors. In reality, it is never possible to obtain absolutely accurate data whether by field or by mail procedures. Errors and biases exist in all data in varying degrees. Considering the nature of the survey, it is reasonable to assume that the data collected by personal interview are freer from biases of reporting and random errors than those obtained by mail.

Table 3 and Chart 1 show the comparative cumulative distribution of workers by earnings obtained by the two methods of collection.

TABLE 3

CUMULATIVE PERCENTAGE DISTRIBUTION OF PRODUCTION WORKERS IN THE RAINWEAR INDUSTRY BY STRAIGHT-TIME AVERAGE HOURLY EARNINGS, UNITED STATES AND SELECTED REGIONS, JUNE 1951, ORIGINAL AND RESURVEY

Average Hourly earnings* (in cents)	United States†		New England	
	Original	Resurvey	Original	Resurvey
Under 75.0	‡	0.4	—	0.7
Under 80.0	13.4	11.8	17.3	18.5
Under 85.0	17.2	16.2	22.4	23.4
Under 90.0	23.7	22.6	27.8	28.4
Under 95.0	28.6	27.7	32.3	32.9
Under 100.0	34.0	34.1	35.4	36.3
Under 105.0	42.2	41.4	43.5	43.6
Under 110.0	48.1	47.0	48.0	48.2
Under 115.0	55.0	54.1	54.0	53.1
Under 120.0	59.2	58.5	58.0	57.6
Under 125.0	63.6	63.3	61.6	61.6
Under 130.0	68.7	67.8	66.1	65.6
Under 135.0	72.4	71.6	68.9	68.6
Under 140.0	74.7	74.1	71.0	70.9
Under 145.0	77.6	77.1	74.3	74.0
Under 150.0	79.8	79.0	76.5	76.1
Under 155.0	82.4	81.4	79.9	79.1
Under 160.0	83.8	82.9	82.0	81.5
Under 165.0	85.6	84.8	84.3	83.9
Under 170.0	86.7	86.0	85.6	85.4
Under 175.0	87.8	87.3	87.0	87.0
Number of workers	9,149	8,929	4,005	3,960
Median rate	$1.10	$1.12	$1.12	$1.12

TABLE 3—(*Continued*)

Average hourly earnings* (in cents)	Middle Atlantic		Great Lakes	
	Original	Resurvey	Original	Resurvey
Under 75.0	‡	0.3	0.1	0.1
Under 80.0	9.2	4.8	10.8	7.1
Under 85.0	12.4	9.0	11.4	9.0
Under 90.0	17.0	13.9	21.5	19.5
Under 95.0	21.5	19.3	27.2	25.4
Under 100.0	24.8	24.0	38.5	38.6
Under 105.0	34.5	32.1	45.6	45.1
Under 110.0	39.0	37.0	55.1	53.4
Under 115.0	43.5	42.2	65.2	65.7
Under 120.0	45.4	44.0	72.2	72.6
Under 125.0	49.0	48.0	79.0	79.4
Under 130.0	55.5	53.1	83.9	84.1
Under 135.0	60.7	58.6	87.7	87.8
Under 140.0	63.6	61.6	89.8	89.9
Under 145.0	66.2	64.8	92.6	92.8
Under 150.0	68.8	66.3	94.8	94.6
Under 155.0	71.5	68.8	95.7	95.6
Under 160.0	72.3	69.9	96.3	96.3
Under 165.0	73.9	71.5	97.4	97.4
Under 170.0	75.6	73.3	97.8	97.7
Under 175.0	76.7	74.8	98.4	98.4
Number of workers	2,276	2,205	2,527	2,432
Median rate	$1.27	$1.27	$1.07	$1.08

* Excludes premium pay for overtime and night work.
† Includes data for other regions in addition to those shown separately.
‡ Less than .05 of 1 per cent.

Beyond the 90-cent point there is little difference in anv of the regional data, and even below this point there is little difference for the industry as a whole. There were slight increases in the median rates, 2 cents in the industry as a whole, 2 cents in the Middle Atlantic region, and 1 cent in the Great Lakes region.

The lower end of a wage distribution is of critical importance for minimum wage determination. Examination of these parts of the distribution shows that the greatest difference occurs in the interval between 75 cents and 80 cents, 11.4 per cent of the workers were classified in this interval in the resurvey as against 13.4 per cent in the original survey. At no other point does the difference exceed 1 percentage point. Inspection of the cumulative distribution (Table 3 and Chart 1) show very small differences beyond the 90-cent point.

The greatest discrepancy between the surveys occurs in two of the

regional distributions. In the Middle Atlantic region, the estimated percentages of workers between 85 cents and 90 cents were 9.2 per cent for the original survey and 4.5 per cent in the resurvey. In the Great

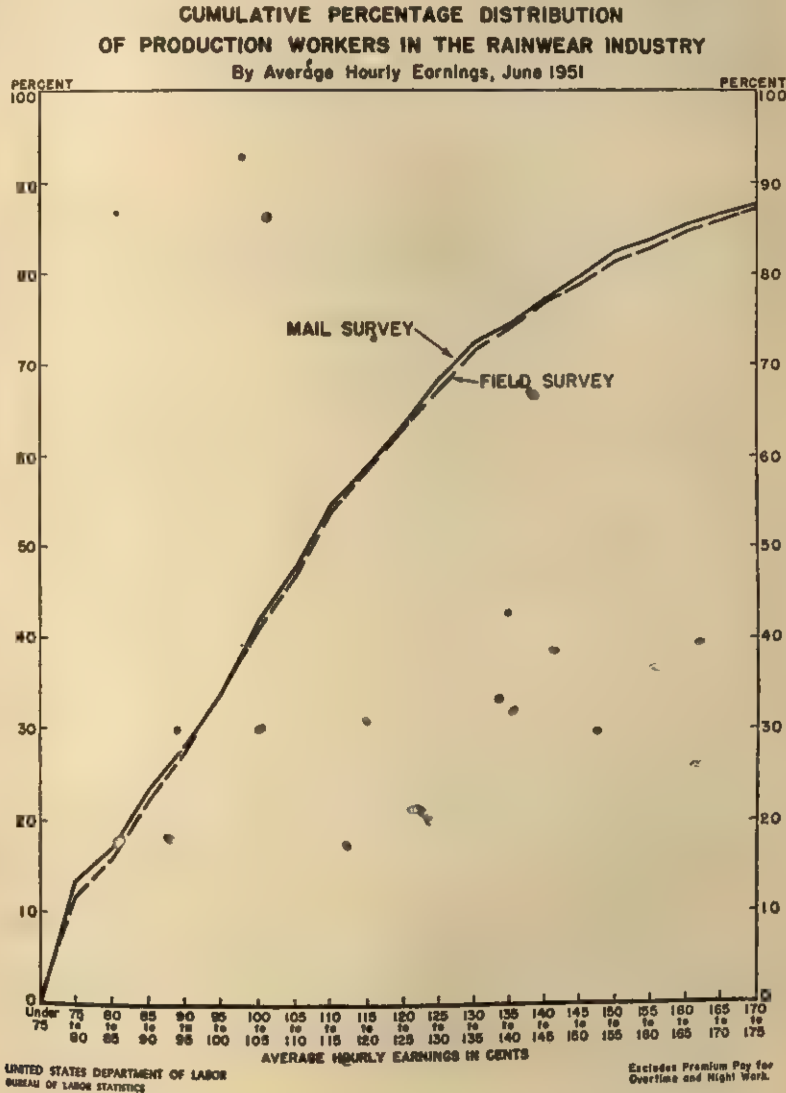


CHART 1

Lakes region the corresponding percentages were 10.7 and 7.0. These differences are due primarily to the two firms reporting large blocks of workers at the guaranteed rate rather than at actual earnings, plus

two other questionable schedules with concentrations of workers at 75 cents in the original survey. (Of these, one firm refused to cooperate in the resurvey, while the other firm supplied incorrect data in the original survey.)

Practically no difference was found in the broad auxiliary and non-auxiliary classifications (See Table 4). It is noteworthy that small declines appeared in the earnings of auxiliary workers, although the median earnings of the non-auxiliary workers increased slightly.

TABLE 4
PERCENTAGE DISTRIBUTION OF AUXILIARY AND NON-AUXILIARY PLANT WORKERS (EXCLUDING LEARNERS) IN THE RAINWEAR INDUSTRY BY STRAIGHT-TIME AVERAGE HOURLY EARNINGS,* FOR THE UNITED STATES, JUNE 1951, ORIGINAL AND RESURVEY

Average hourly earnings* (in cents)	Auxiliary Workers		Non-Auxiliary Workers	
	Original Survey	Resurvey	Original Survey	Resurvey
Under 75.0	0.1	0.3	—	0.4
75.0 and under 80.0	36.3	41.8	9.6	7.7
80.0 and under 85.0	10.5	10.7	3.1	3.6
85.0 and under 90.0	11.9	14.7	5.8	5.4
90.0 and under 95.0	5.8	7.5	4.8	4.8
95.0 and under 100.0	4.1	5.1	5.6	6.5
100.0 and under 105.0	8.9	6.3	8.1	7.4
105.0 and under 110.0	3.6	2.4	6.1	6.0
110.0 and under 115.0	3.0	1.4	7.3	7.8
115.0 and under 120.0	2.7	3.9	4.4	4.5
120.0 and under 125.0	3.1	2.5	4.8	5.0
125.0 and under 130.0	2.0	1.2	5.5	4.9
130.0 and under 135.0	1.3	.8	4.2	4.2
135.0 and under 140.0	1.1	.4	2.6	2.7
140.0 and under 145.0	.7	.3	3.3	3.4
145.0 and under 150.0	3.2	.3	2.3	2.1
150.0 and under 155.0	1.0	—	2.8	2.7
155.0 and under 160.0	.3	.2	1.5	1.7
160.0 and under 165.0	.5	.1	2.0	2.1
165.0 and under 170.0	.1	—	1.3	1.4
170.0 and under 175.0	.2	—	1.2	1.5
175.0 and over	.6	.1	13.7	14.2
Total	100.0	100.0	100.0	100.0
Number of workers	1,027	990	8,051	7,939
Median rate	$0.85	$0.84	$1.14	$1.15

* Excludes premium pay for overtime and night work.

REPORTING ERRORS OF FIRMS IN SCOPE

Establishments revisited gave a variety of explanations as to how the original data were compiled. Of the 53 firms studied, 25 followed the instructions completely and correctly; a few included salesmen and office workers; some estimated earnings; a few included overtime and 10 simply could not recall what was done. The most serious downward

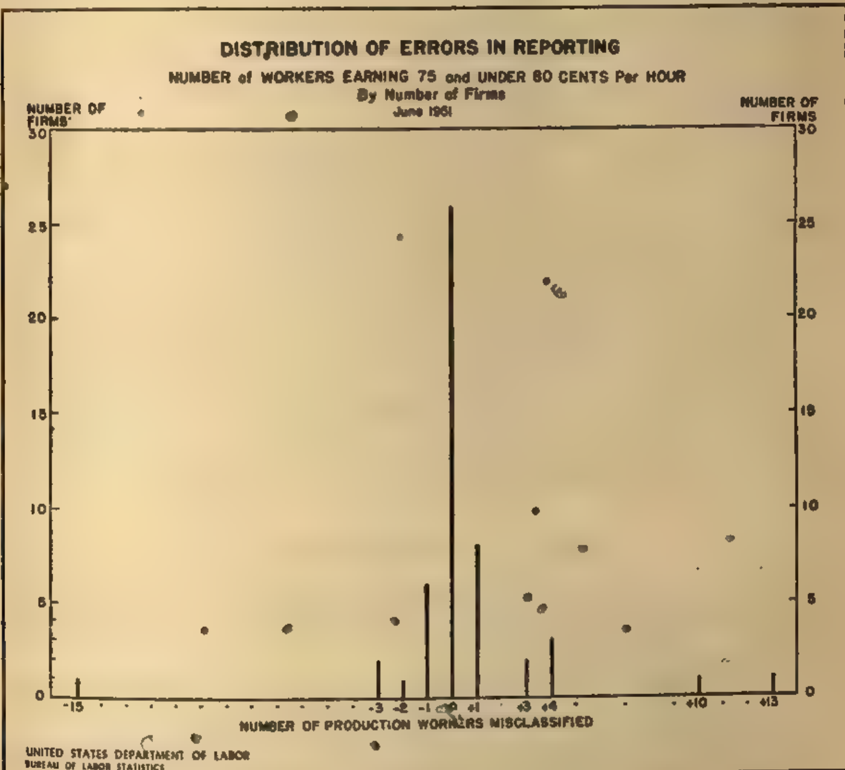


CHART 2

bias resulted from two firms which reported guaranteed rates instead of actual earned rates. The main difference between the original survey and the resurvey is due to the errors of these two firms and to the omission of the one firm with 100 employees that refused to cooperate in the resurvey.

A comparison was made of changes in reporting in the 75 and under 80 cent groups (omitting the two firms which misinterpreted the instructions.)

A frequency distribution of the errors was constructed (Chart 2). The X variable was the number of production workers misclassified in

the 75 and under 80 cent class. Positive errors were workers erroneously included in the class, and negative errors were those who should have been included. The Y variable was the number of firms which misclassified workers.

This frequency distribution of reporting errors showed an arithmetic mean of -0.4. Thus, the critical class was less than half a worker per firm too large in the original survey. In order to see whether the error in this class is of a compensating nature, the degree of asymmetry was established using Pearson's measure of skewness which equaled 0.05. Thus, the error distribution was almost symmetrical and the errors are to a great extent compensating.

A regional pattern in errors of reporting is noticeable. The closest correspondence between the two forms of collection is found in the New England region. For the bulk of the firms agreement was practically perfect at nearly every point in the distribution. No mail questionnaire was of such quality as would necessitate rejection. The poorest mail reporting came from the Middle Atlantic region where 5 of the 22 schedules collected by mail differed so much from the field resurvey that they could not be termed usable. Of the 10 comparisons in the rest of the country only 1 mail questionnaire was unsatisfactory. Only one of the 13 schedules obtained by the ILGWU resurvey was sufficiently different from the original mail schedule to warrant rejection of the latter.

SUMMARY AND CONCLUSIONS

1. The data indicate that collection of wage distributions by mail questionnaire is feasible in piece rate industries and is also feasible in industries paying hourly rates.

2. No consistent reporting bias was evidenced. Most of the reporting errors were not only compensatory, but were generally small.

3. The major reporting error in such surveys is the confusing of earned and guaranteed rates. Care must be taken in design of the questionnaire and in the accompanying instructions to distinguish between those two items. Careful editing will frequently disclose whether large concentrations of workers at any one rate are caused by reporting guaranteed rates instead of earnings. Personal visit may be necessary to clarify such cases.

4. A field follow-up of respondents classifying themselves as out of scope is necessary in an industry characterized by frequent changes in product. Without such a field follow-up an accurate estimate of the universe is difficult, and may result in overrepresentation of some segments of the industry in the survey totals.

INDUSTRIAL CLASSES IN THE UNITED STATES 1870 TO 1950

Tillman M. Sogge
St. Olaf College

In an earlier article the writer classified the gainful workers in the United States for 1870 to 1930, and the labor force in 1940 into industrial classes.[1] This paper brings this classification up to date by adding the data for 1950.

The fact that the occupational classifications used in the Census of 1950 and the Census of 1940 were quite comparable made the task of adding the data for 1950 to the previous compilations comparatively easy.[2]

Some interesting developments have taken place during the decade of the 1940's, a period during which our economy was affected very markedly by war and postwar changes. In the main these changes have not altered the direction of trends indicated in earlier decades but the degree of change has in some instances been increased. Some of the more important changes from 1940 to 1950 are:

1. The number of persons for whom occupations were reported increased from 50,737,284 in 1940 to 57,632,879 or an increase of 13.5 per cent which is a considerably greater rate of increase than during the depression decade of the thirties.
2. The number of persons engaged in farming occupations continued to decline in spite of the large increase in the total labor force. Table I indicates that the number of farm laborers has declined from 6,143,998 in 1910 to 2,497,637 in 1950 or a decline of over 59 per cent in 40 years. The decline in relative importance of agricultural occupations is clearly evidenced by a decrease from 17.4 per cent of the total in 1940 to 11.8 per cent of the total in 1950.
3. Both proprietors and officials and professional classes have shown a significant increase in numbers during the past decade, but the greatest increase in numbers is indicated in the lower salaried

[1] Tillman M. Sogge, "Industrial Classes in the United States in 1940," *Journal of the American Statistical Association*, 39 (1944), 516–18. Parallel data for the years 1870 through 1920 were first presented by Professor Alvin H. Hansen in the two following articles (1) "Industrial Class Alignments in the United States," *Journal of the American Statistical Association*, 17 (1920), 417–25, and (2) "Industrial Classes in the United States in 1920," *Journal of the American Statistical Association*, 18 (1922), 503–06. The data for 1930 were first presented in an article by the writer entitled, "Industrial Classes in the United States in 1930," *Journal of the American Statistical Association*, 28 (1933), 199–203.

[2] Coverage of the different classes has been described in the earlier articles in this series listed in the footnote above.

TABLE I

INDUSTRIAL CLASSES

	1870	1880	1890	1900	1910	1920	1930	1940	1950*
Farm laborers	2,885,996	3,323,876	3,004,061	4,410,877	6,143,998	4,178,637	4,392,764	3,505,275	2,497,637
Farmers	3,000,229	4,282,074	5,370,181	5,770,738	6,229,161	6,463,708	6,079,234	5,238,049	4,337,719
Proprietors and officials	581,378	807,049	1,347,329	1,811,715	2,879,023	3,168,418	4,270,856	4,197,207	5,046,577
Professional	414,708	666,338	1,114,507	1,565,686	2,074,792	2,760,190	3,845,559	4,454,381	5,968,173
Lower salaried	309,413	529,473	965,852	1,329,928	2,393,620	3,985,306	7,116,814	8,071,201	11,253,501
Servants	975,734	1,075,655	1,454,791	1,453,677	1,572,22[illegible]	1,270,946	1,999,133	2,840,021	1,414,732
Industrial wage earners	8,328,351	5,286,829	7,3[illegible],442	10,263,569	14,556,979	17,648,072	18,512,640	19,957,220	24,761,214
Unclassified	1,010,114	1,420,795	2,118,498	2,467,043	2,317,538	2,138,971	2,612,920	2,383,930	2,353,326
Total	12,505,923	17,392,099	22,735,661	29,073,233	38,167,336	41,614,248	48,829,920	50,737,284	57,632,879

TABLE II

INDUSTRIAL CLASSES

(Per cent)

	1870	1880	1890	1900	1910	1920	1930	1940	1950*
Farm laborers	23.1	19.1	13.2	15.2	16.1	10.0	9.0	6.9	4.3
Farmers	24.0	24.6	23.6	19.8	16.3	15.5	12.4	10.5	7.5
Proprietors and officials	4.6	4.6	5.9	6.2	7.5	7.6	8.7	8.3	8.8
Professional	3.3	3.8	4.9	5.4	5.4	6.6	7.9	8.8	10.4
Lower salaried	2.5	3.0	4.3	4.6	6.3	9.6	14.6	15.9	19.5
Servants	7.8	6.2	6.4	5.0	4.1	3.1	4.1	5.6	2.5
Industrial wage earners	26.6	30.4	32.4	35.3	38.2	42.4	37.9	39.3	42.9
Unclassified	8.1	8.2	9.3	8.5	6.0	5.1	5.4	4.7	4.1

* These figures do not include 1,366,064 experienced persons in the labor force for whom occupations were not reported. The data were compiled from the 1950 Census of Population, Volume II Characteristics of the Population: Part I, United States Summary, Table 124, pp. 1-261 to 1-266.

group and in industrial wage earners. The increase of nearly five million industrial wage earners during the past decade gives evidence of the expansion in our national economy which has taken place. Even though there has been this large increase in the number of industrial wage earners during the past decade, and an increase in relative importance of this group from 39.3 to 42.9 per cent of the total, it is significant to note that the importance of the industrial wage earners in relation to all the non-agricultural classes combined was less in 1950 than in 1920.

4. The number of servants in 1950 was less than half the number in 1940 dropping from 2,840,021 to 1,414,732. Percentagewise, servants constituted 5.6 per cent of the total in 1940 and only 2.5 per cent of the total in 1950.

STATISTICAL METHODS FOR POISSON PROCESSES AND EXPONENTIAL POPULATIONS

ALLAN BIRNBAUM*
Columbia University

1. INTRODUCTION

THE research worker and statistician frequently deal with phenomena in which events of some type occur randomly in time, or in which particles are randomly distributed in space. The Poisson process is the formal model of such phenomena. Furthermore, many phenomena which may naturally be represented by use of the exponential distribution or the Poisson distribution can alternatively be represented as Poisson processes and dealt with advantageously in this form. Statistical methods for the study of such phenomena can be as flexible and yet simple as the Poisson process model itself. A number of such statistical methods are described and illustrated below. Some of these methods were developed recently, while others are well-known but are described here briefly for comparison and completeness.

2. THE POISSON PROCESS

The results of any experiment in which observation is performed continuously and "events" (i.e., occurrences of any specified kind) are tallied, can always be described by a function $x=x(t)$, which gives the number of events observed, x, during the first t units of observation, for all values of t from 0 through T, the total amount of observation performed. Such an experiment, yielding an observed function $x(t)$, is a Poisson process if the events occur randomly in the sense of the following natural definition: given that any number x of events are observed in any amount t of observation, the points of occurrence of the x events are randomly (i.e., independently uniformly) distributed between 0 and t. Examples in which the Poisson process is a very accurate and useful model are the following:

Example 1. When a Geiger counter is used in an essentially constant environment, the counter tallies the number of times it is hit by radioactive particles. Here the amount of observation performed is measured in units of time during which the counter is operated. The Poisson process represents this phenomenon quite accurately. (The assumptions for the Poisson process will be violated seriously under extreme conditions: In an environment of very high radioactivity, the

* Research supported by the Office of Naval Research.

counter's inability to record a hit extremely soon after an earlier hit will seriously violate the independence assumption. If the environment contains radioactive materials with extremely rapid rates of decay, the assumptions of equal probabilities will be seriously violated.)

Example 2. If any homogeneous or well mixed material containing practically infinitesimal foreign particles or flaws is inspected, the amount of observation will be measured by the volume (or area or length) of material inspected, and the "events" will be the particles or flaws observed. One case is the use of a haemocytometer to count the number of blood cells distributed over an area on a slide. (For studies of the magnitude of deviations from the independence assumption due to "crowding" of cells, see [1] and the references given there.) Another case is that of inspection of manufactured materials by length or area for defects, if the manufacturing process is such that there is no appreciable dependence in the locations of defects and the average defect rate does not change appreciably during the period of inspection. Other applications, and some statistical methods discussed below, will be found in [18].

A Poisson process can be characterized in the following two simple alternative and equivalent ways:

(a) The "waiting times" u between successive events are independently distributed with the exponential density function

$$g(u) = (1/\theta)e^{-u/\theta} \qquad \text{for } u \geqq 0.$$

Here θ is the mean of waiting times:

$$E(u) = \theta.$$

(b) The increment $y = x(t_2) - x(t_1)$ of $x(t)$ on any interval of length $d = t_2 - t_1$ has the Poisson distribution

$$p(y) = e^{-d\lambda}\frac{(d\lambda)^y}{y!}, \qquad y = 0, 1, \cdots.$$

and the increments of $x(t)$ on non-overlapping intervals are independent. Here $d\lambda$ is the mean increment on an interval of length d. Thus λ is the mean rate of occurrences, and $\lambda = 1/\theta$.

All statistical questions concerning Poisson processes involve inferences about the value of the single parameter λ or θ of one process, or about the values of the respective parameters of several processes. (Methods for determining whether a given process is Poisson will not be discussed here; Davis [6] describes and applies several methods.)

The flexibility and simplicity of statistical methods applicable to

Poisson processes seems to stem from two sources. One of these is the simple way in which information from different experiments or observations on a single process can be combined. As an illustration, consider the life-testing procedures developed by Epstein and Sobel [9, 10]. It is assumed here that the lengths of life of a type of electron tube, say, are exponentially distributed with unknown mean life θ. (For some empirical tests of this assumption, see [6].) If one tube is observed until failure, then replaced by a new tube which is again replaced upon its failure, and so on, we have a sequence of observed lengths of life u from an exponential distribution. Hence the number of failures observed by any time t is a Poisson process $x(t)$. Now suppose that we place any number of such tubes under observation, and add new tubes or remove tubes from observation in a quite arbitrary manner. The information given by such an experiment is fully described as a single Poisson process, where t is now taken to be the number of tube-hours of life observed at any point in the procedure, and $x(t)$ is the number of failures observed up to that point. The feature illustrated here is called the additivity property of the Poisson process.

The second source of flexibility of statistical methods is the fact that we can shift back and forth freely between use of the two simple distributions which characterize the Poisson process, the exponential and the Poisson distributions.

3. ESTIMATION OF λ OR θ

As a basic example of the flexibility possible in the statistical treatment of data from such procedures, consider the problem of estimating the parameter λ of a Poisson process by means of a confidence interval.

Method 1. If a fixed amount t of observation is performed, then the number x of events observed has the Poisson distribution with mean $\mu = \lambda t$,

$$\Pr(x) = e^{-\mu}\frac{\mu^x}{x!}, \qquad \text{for } x = 0, 1, 2, \cdots.$$

By standard methods we can construct a confidence interval for μ, at a desired confidence level. An exact construction can be based on tables of the Poisson distribution, and extensive tables of such confidence intervals as well as tables of the Poisson distribution are available in [13, 17, 20].

Method 2. If observation of a Poisson process is continued until a specified number m of events has been observed, and if T is the amount of observation actually required to observe m events, then $2\lambda T$ has the

chi-square distribution with $2m$ degrees of freedom. A confidence interval for λ at a desired confidence level can then be constructed in the standard way, as described below.

It will be useful here and also in connection with other statistical problems discussed in the following sections to note that statistical problems dealing with variances of normal populations have direct analogues in problems dealing with parameters of Poisson processes. The analogue of the present problem is that of constructing a confidence interval for the variance σ^2 of a normal population given S^2, the sum of squared deviations from the known or estimated mean, with $2m$ degrees of freedom; the confidence interval is constructed by using the fact that S^2/σ^2 has the chi-square distribution with $2m$ degrees of freedom. Thus $2T/\theta$ and S^2/σ^2 have the same distributions, and so do the statistics $2T$ and S^2 if the corresponding parameters θ and σ^2 are equal. Here and in later problems we can take advantage of this correspondence to apply statistical methods developed originally primarily for a problem dealing with normal variances to a corresponding problem dealing with Poisson processes, and vice-versa.

To construct the desired confidence interval for λ, we select from tables of the chi-square distribution with $2m$ degrees of freedom, values C and D such that the probabilities of values less than C, and greater than D, are each $\alpha/2$. Then we can write

$$1 - \alpha = \text{Prob}\left\{C \leqq 2\lambda T \leqq D\right\} = \text{Prob}\left\{\frac{C}{2T} \leqq \lambda \leqq \frac{D}{2T}\right\}.$$

Thus $(C/2T, D/2T)$ is a confidence interval for λ, with confidence coefficient $1-\alpha$. The small bias of such confidence intervals could be removed at the price of complicating the procedure.

The following method of constructing point estimates of λ based on Method 2 was developed in [15]. We have with probability $1-\alpha$ that

$$C/2T < \lambda < D/2T.$$

Given this inequality, the maximum percentage deviation of λ from an estimate λ' is minimized by taking

$$\lambda' = \frac{C + D}{4T}.$$

This maximum percentage deviation is

$$\tau = \frac{D - C}{C + D}\,100\%.$$

τ is a quantity which decreases as the number m of events observed increases. Hence we can use tables of the chi-square distribution to determine the number m of events to be observed to give an estimate λ' with maximum percentage deviation not exceeding any given positive τ, at confidence level $1-\alpha$.

On the other hand, when Method 1 is used, there is no way of prescribing the amount t of observation so as to obtain point estimates of bounded percentage error. As shown in most statistics texts, an approximate construction of the confidence intervals of Method 1 is given, say at the .95 confidence level, by the interval

$$\frac{x}{t}-\frac{1.96\sqrt{x}}{t}, \qquad \frac{x}{t}+\frac{1.96\sqrt{x}}{t}.$$

It is seen that here the maximum percentage deviation of λ from the estimate x/t is $(1.96/\sqrt{x})100\%$, a function of the observed value of x.

Method 3. In Method 1, the absolute deviation of the estimate x/t from λ is bounded by $C\alpha(\sqrt{x}/t)$, where $C\alpha$ is a constant corresponding to the confidence level used, with probability $1-\alpha$. In Method 2 the corresponding bound is $\lambda'\tau/100\%$. In neither case can the magnitude of this bound be prescribed in planning the amount of observation t or m, since each bound is determined by the outcome of the experiment. The following method gives estimates of prescribed absolute precision.

The problem here is to give an estimate λ' of λ such that with probability at least $1-\alpha$, $|\lambda'-\lambda|\leqq\epsilon$, where α and ϵ are given positive constants. Let n be a positive integer. Observe T_n, the waiting time required for the occurrence of n events. Let $c=\alpha\epsilon^2/2n$. Perform additional observation of the process for $1/2cT_n$ units of time; let X be the number of events observed in this period. Let $\lambda'=2cT_nX$. Then λ' is an estimate meeting the stated requirements.

The stated properties of λ' may be demonstrated by verifying that λ' has expected value λ and variance $2nc$. Then Tchebycheff's inequality gives

$$\text{Prob}\ \{\,|\lambda'-\lambda|\leqq\epsilon\}\geqq 1-\frac{2nc}{\epsilon^2}=1-\alpha.$$

It remains for further investigation to determine the largest constant c' which can be used to determine the additional amount $1/2c'T_n$ of observation without destroying the desired properties of the estimates $2c'T_nX$. Likewise, rules for the optimal choice of n remain to be investigated.

4. TESTS OF HYPOTHESES ON λ OR θ

A test of a hypothesis specifying the value of λ (and hence of $\theta = 1/\lambda$), say the hypothesis H_0: $\lambda = \lambda_0$, is provided by each method of constructing a confidence interval for λ. If A and B are confidence limits for λ constructed by one of the methods of the preceding sections at confidence level $1-\alpha$, then a test of H_0 at the $1-\alpha$ significance level is obtained by accepting H_0 if $A \leqq \lambda_0 \leqq B$ and otherwise rejecting H_0. Tests of H_0 against one-sided alternatives (e.g., $\lambda > \lambda_0$) can be based on corresponding one-sided confidence limits for λ. Tests with prescribed power can be designed by use of the Poisson or chi-square distributions. As these procedures are relatively standard, they will not be discussed further here.

In each of these testing methods, an economy in the amount of observation required can be achieved by taking advantage of the fact that observation of the Poisson process can be performed continuously. For example, consider tests of H_0: $\lambda = \lambda_0$ against the one-sided alternative $\lambda > \lambda_0$. In the type of test based on a prescribed amount t of observation, the test procedure will amount to rejecting H_0 if the observed number of events x exceeds some "critical number" x_α. Hence if $x_\alpha + 1$ events have been observed at any point in the procedure, observation can be terminated and H_0 rejected at once. In the type of test based on observation of a prescribed number m of events, the test procedure will amount to accepting H_0 if the required amount T of observation is at least equal to some "critical amount" T_α. Hence if m events are not observed before $t = T_\alpha$, observation can then be terminated and H_0 accepted at once, while if m events are observed at or before $t = T_\alpha$, observation can then be terminated and H_0 rejected at once.

Another type of test procedure (which has been investigated by two groups of statisticians, in [8] and [9]) achieves optimal economy in the required amount of observation at the price of a slightly more complicated statistical procedure. Such tests are now available only for testing H_0 against one-sided alternatives. (Two-sided tests of this type can in principle be constructed, by the method used by Wald in [23], pp. 134–37, but this has not yet been done because of certain technical difficulties involved.) These tests can be described as typical sequential probability ratio tests of the kind developed by Wald in [23] and [24], with account and advantage being taken of the fact that the Poisson process may be observed continuously. The typical form of these procedures is the following: Continue observation only so long as the number x of events observed and the amount t of observation performed satisfy the inequality

$$b + st < x < a + st.$$

As soon as $x \leqq b+st$, accept H_0; as soon as $x \geqq a+st$, accept H_1: $\lambda = \lambda_0 + \Delta$. Application of the method may be simplified by a graphical representation of the test inequalities in the (t, x) plane. Here

$$s = \Delta \log\left(\frac{\lambda_0}{\lambda_0 + \Delta}\right), \qquad a = \left(\log\frac{1-\beta}{\alpha}\right)\left(\log\frac{\lambda_0}{\lambda_0 + \Delta}\right),$$

and

$$b = \left(\log\frac{\beta}{1-\alpha}\right)\left(\log\frac{\lambda_0}{\lambda_0 + \Delta}\right),$$

where α is the desired maximum probability of accepting H_1 when $\lambda \leqq \lambda_0$, and β is the desired maximum probability of accepting H_0 when $\lambda \geqq \lambda_0 + \Delta$.

In [10] a number of life testing procedures using Method 2 were investigated, and their operating characteristics and required amounts of test equipment and time determined.

5. COMPARING TWO POISSON PROCESSES

Comparisons of two Poisson processes with parameters λ_1, λ_2 respectively are expressed most usefully in many applications by statements about the value of the ratio $\gamma = \lambda_2/\lambda_1$. The methods described in the following Section 5A provide estimates and tests on γ. In other applications, it may be more appropriate to express comparisons of two processes by statements about the value of the difference $\Delta = \lambda_2 - \lambda_1$; the rather restricted methods available for comparisons in terms of Δ are described in Section 5B. The following examples illustrate possible reasons for choosing either γ or Δ as a criterion of comparison.

Example 1. In order to measure the effectiveness of a shield intended for protection against radiation, we may take any steady source of radiation and a Geiger counter in fixed positions, and record "hits" by radiation alternately with the shield interposed between counter and source and with the shield removed. The natural measure of effectiveness of such a shield is the percentage of radiation it eliminates. If λ_2 is the intensity of radiation at the counter when the shield is present, and λ_1 is the intensity when the shield is removed, then the effectiveness of the shield is $100[1-(\lambda_2/\lambda_1)]\%$. A confidence interval estimate of $\gamma = \lambda_2/\lambda_1$ provides statistical information in an appropriate form, such as, for example, "At confidence level .99, we estimate the effectiveness of the shield to be at least 98.4%."

Example 2. In order to decide which of two samples of ore shows greater radioactivity, as a basis for selecting one of two lots of ore for refining, we can record radiation from each sample with a Geiger counter. If λ_1, λ_2 are the respective intensities of radiation at the counter, the degree to which one lot is preferable to the other is represented by $\Delta=\lambda_2-\lambda_1$, for the gain to be achieved by a correct choice of a better lot is proportional to $|\Delta|$. An appropriate statistical procedure is a test of the hypothesis H_1, that $\lambda_1>\lambda_2$, against the alternative $\lambda_2>\lambda_1$, satisfying the requirement that if $|\Delta|\geqq\Delta'$, where Δ' is the smallest value of $|\Delta|$ of practical significance, then with probability at least $1-\alpha$ the better lot will be chosen.

5A. COMPARISONS IN TERMS OF $\gamma=\lambda_2/\lambda_1$

Method 1. If a prescribed amount t of observation is performed on each of two Poisson processes with parameters λ_1, λ_2, then the numbers x_1, x_2 of events observed in the respective processes will have Poisson distributions with means $\mu_1=\lambda_1 t$, $\mu_2=\lambda_2 t$. Then $\mu_2/\mu_1=\lambda_2/\lambda_1=\gamma$. It has been shown in [19] and [22] that all information on the value of γ contained in the observations x_1, x_2 is obtained by treating the observations as a binomial sample of $m=x_1+x_2$ observations in which x_1 "successes" are observed, and the probability of a "success" is

$$\rho=\frac{\lambda_1}{\lambda_1+\lambda_2}=\frac{1}{1+\lambda_2/\lambda_1}=\frac{1}{1+\gamma}.$$

Thus to test a hypothesis on the value of γ, say H_0: $\gamma=1$ (i.e. $\lambda_1=\lambda_2$), we consider the corresponding test of H_0':

$$\rho=1/2\left(\text{as } \rho=\frac{1}{1+\gamma}=\frac{1}{1+1}\right)$$

when x_1 successes are observed in m binomial trials. We reject H_0 if and only if H_0' is rejected, and at the same significance level. Again, to construct a confidence interval for γ, we can first construct one for ρ (as described in [4]) from our sample of x_1 successes in m trials. If we obtain, for example, "at the .95 confidence level, we estimate that $.33\leqq\rho\leqq.50$," then we can use the equation $\rho=1/(1+\gamma)$, or $\gamma=(1/\rho)-1$, to obtain the equivalent "at the .95 confidence level, we estimate that $1\leqq\gamma\leqq2$."

Method 2. An important weakness of Method 1 is that the procedure cannot be planned to provide any minimum amount of information on γ if the values of λ_1, λ_2 are completely unknown. As we have seen, the

method provides information equivalent to a binomial sample of $m=x_1+x_2$ observations, but m itself is a random observation and may in any particular instance be too small to provide sufficient information (it may even equal zero, and provide no information about γ). Actually m has a Poisson distribution with mean $\mu_1+\mu_2$ so that, if the unknown values of both μ_1 and μ_2 are very small, Method 1 will very often give very little information about γ; even if $\mu_1+\mu_2$ is quite large, there will occur some samples with m too small to give sufficient information about γ.

One way out of this difficulty is apparent. If at least m' binomial observations are required to provide enough information about ρ (and hence about γ) in some application, and if an application of Method 1 gives $m<m'$, another application of Method 1 with an additional amount t_2 of observation of each process may be possible, and if necessary additional amounts of observation until the total number of events observed reaches or exceeds m'. Then the usual binomial statistical methods can be applied to the totals of events observed in each process, as in Method 1.

Method 3. If two Poisson processes can be observed continuously, with observation on each performed at the same rate, then the probability that the ith event observed will occur in the first process is $\rho=1/(1+\gamma)$, for each $i=1, 2, \cdots$. Such observation thus provides a sequence of binomial observations of indefinite length, to which the statistician may apply any of the sequential or curtailed-sampling methods for binomial data (e.g. Wald [23]), or the statistician may terminate observation when a prescribed number n of events has occurred, and apply non-sequential methods for binomial samples of size n. (Methods 2 and 3 were developed in [2].)

Method 4. An alternative way of prescribing observation of two Poisson processes is to require that the first be observed until n_1 events have occurred, and the second until n_2 events have occurred in it. If T_1 is the amount of observation of the first process required to observe n_1 events, then as we have seen above $2\lambda_1 T_1$ has the chi-square distribution with $2n_1$ degrees of freedom; the corresponding quantity for the second process, $2\lambda_2 T_2$, has the chi-square distribution with $2n_2$ degrees of freedom.

The present problem is statistically equivalent to that of comparing variances of two normal populations, given sample variances with $2n_1$ and $2n_2$ degrees of freedom respectively. To test whether $\lambda_1=\lambda_2$, we compute $F=n_2T_1/n_1T_2$, and use the fact that F has the F-distribution with $2n_1$, $2n_2$ degrees of freedom when $\lambda_1=\lambda_2$. To give a confidence in-

terval for $\gamma=\lambda_2/\lambda_1$, we may use the standard method developed for the corresponding problem of giving a confidence interval for the ratio σ_1^2/σ_2^2 of two normal variances by use of the F-distribution. These methods were developed in [11] and [21]. (Cox [5] has investigated the effects of applying to a Method 1 experiment the simpler analysis of Method 4.)

In some applications the present procedure would be more easily applied than those of Methods 2 and 3 above; for example if the processes are separated in space or time the present procedure might be applied. If simultaneous observation of the processes is possible, Methods 2 and 3 would generally give earlier termination since in Method 4 if λ_2 is much smaller than λ_1, T_2 will generally be much larger than T_1.

5B. COMPARISONS IN TERMS OF $\Delta=\lambda_2-\lambda_1$

Method 1. Let two Poisson processes with parameters λ_1, λ_2 be observed continuously so that the respective waiting times for the first event in each process, T_1 and U_1 say, are obtained, then the waiting times T_2, U_2 for the second event in each process, and so on. We thus obtain a sequence of pairs $(T_i,\ U_i)$ of waiting times, for each $i=1,\ 2, \cdots$; the T_i's have an exponential distribution with mean $\theta_1=1/\lambda_1$, the U_i's an exponential distribution with mean $\theta_2=1/\lambda_2$. Using such observations we can solve the testing problem stated in Example 2 above.

To do this we apply Girshick's [14] method to the problem of ranking the two exponential population with respect to their means; the resulting test procedure is as follows: After each pair $(T_i,\ U_i)$ of observed waiting times is obtained, compute

$$Z = \Delta\sum_{i=1}^{m} (T_i - U_i).$$

Continue taking observations until either $Z\geqq a$, in which case accept the hypothesis H_1: $\lambda_2-\lambda_1=(1/\theta_2)-(1/\theta_1)\geqq\Delta$, or until $Z\leqq b$, in which case accept the hypothesis H_0: $\lambda_1-\lambda_2=(1/\theta_1)-(1/\theta_2)\geqq\Delta$. Here $a=\log (1-\beta)/\alpha$, $b=\log\ (\beta/1-\alpha)$, where α is the desired maximum probability of rejecting H_0 when H_0 is true, and β the desired maximum probability of rejecting H_1 when H_1 is true. It has been shown in [3] that, unfortunately, Girshick's method can not be generalized to deal with problems like testing whether $\lambda_1-\lambda_2\equiv 1/\theta_1-(1/\theta_2)\doteq 0$ against the alternative $\lambda_2-\lambda_1\equiv 1/\theta_2-(1/\theta_1)=\Delta$.

Method 2. Consider the problem of constructing an estimate Δ^* of $\Delta=\lambda_2-\lambda_1$, where λ_1, λ_2 are the unknown parameters of two Poisson

processes, such that with probability at least $1-\beta\,|\Delta^*-\Delta|\leqq\eta$, where β and η are given positive constants. One solution is given by setting $\Delta^*=\lambda_2^*-\lambda_1^*$, where λ_1^* and λ_2^* are obtained as in Method 3 of Section 3 above, taking $\epsilon=\eta/2$ and $(1-\alpha)^2=1-\beta$. In case the two processes can be observed simultaneously, a more efficient solution is the following: Let m be a positive integer. Observe T_m, the waiting time required for the occurrence of a total of m events when the two processes are observed simultaneously. Let $d=\beta\eta^2/2m$. Perform additional observation of each process for $\frac{1}{2}dT_m$ units of time; let Y_1, Y_2 be the respective numbers of events observed in the two processes in this period. Let $\Delta''=2dT_m(Y_2-Y_1)$. Then Δ'' is an estimate meeting the stated requirements. The estimates Δ^* and Δ'' provide useful simple solutions of the common Geiger counter problem of estimating the difference between "noise" count rate and "source plus noise" count rate.

The stated properties of Δ'' are verified as follows: $E(\Delta'')=\Delta\cdot$ $E(\Delta''-\Delta)^2=2md$. Then by Tchebycheff's inequality

$$\text{Prob. }\{|\Delta''-\Delta|\leqq\eta\}\geqq 1-2md/\eta^2=1-\beta.$$

Further investigation should provide rules for optimal choice of n or m, and for shorter periods of additional observation which will suffice to meet the stated requirements.

6. COMPARING THREE OR MORE POISSON PROCESSES

Generalizations of two of the preceding methods for comparing two Poisson processes are available for comparing any number k of processes.

The generalization of Method 4 of Section 5A is the following: Let T_i be the amount of observation required for the observation of a preassigned number n_i of events in the ith process, for $i=1,\cdots k$. Then $2\lambda_iT_i$ has the chi-square distribution with $2n_i$ degrees of freedom, where λ_i is the parameter of the ith process, for each i. Thus this problem is statistically equivalent to that of comparing variances of k normal populations on the basis of sample variances based respectively on $2n_1,\cdots 2n_k$ degrees of freedom. For the problem of testing whether k normal variances are equal, tables of critical values are available in [16]. Let us denote by S_i^2 the sum of squares for the ith normal population, based on $2n_i$ degrees of freedom, used in this test procedure. Then if we substitute for each S_i^2 the observed value T_i, for $i=1,\cdots,k$, and then proceed formally with this test procedure, we obtain a test of the hypothesis that the k Poisson parameters are equal; this test has

the same significance level as the original test, and a corresponding power function.

Every procedure for comparing normal variances can be similarly adapted to give a corresponding procedure for comparing Poisson process parameters.

Method 3 of Section 5A can be generalized as follows: Let k Poisson processes with respective parameters $\lambda_1, \cdots, \lambda_k$ be observed simultaneously. Then the probability that the j^{th} event observed will be observed in the i^{th} process is

$$\rho_i = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_k}, \qquad \text{for each } i.$$

Thus the observed events can be considered independent observations from a multinomial population with class proportions $\rho_1, \cdots, \rho_k$, and all methods for dealing with multinomial data can be applied. For example, the hypothesis that the k processes have equal parameters is equivalent to the hypothesis that the corresponding proportions ρ_i are equal: $H_0: \rho_1 = \cdots = \rho_k = 1/k$. A chi-square test of this hypothesis may be carried out in a $k \times 2$ table; the chi-square statistic for this problem is generally termed the "index of dispersion." The limitations of Method 1 described above apply also to this procedure when a preassigned amount t of observation is used.

Further theoretical discussion of Poisson processes is given in [7] and [12].

REFERENCES

[1] Berkson, Joseph, Magath, Thomas B., and Hurn, Margaret, "Laboratory standards in relation to chance fluctuations of the erythrocyte count as estimated with the hemocytometer," *Journal of the American Statistical Association*, 30 (1935), 414–26.

[2] Birnbaum, Allan, "Some procedures for comparing Poisson processes or populations," *Biometrika*, 45 (1953), 447–49.

[3] Birnbaum, Allan, "Some sequential tests for comparing populations," submitted for publication.

[4] Clopper, C. J., and Pearson, E. S., "The use of confidence or fiducial belts illustrated in the case of the binomial," *Biometrika*, 26 (1934), 403–14.

[5] Cox, D. R., "Some simple approximate tests for Poisson variates," *Biometrika*, 40 (1953), 354–60.

[6] Davis, D. J., "An analysis of some failure data," *Journal of the American Statistical Association*, 47 (1952), 113–50.

[7] Doob, J. L., *Stochastic Processes*, New York, John Wiley and Sons (1953), 98, 398–407.

[8] Dvoretzky, A., Kiefer, J., and Wolfowitz, J., "Sequential decision problems

for processes with continuous time parameter," *Annals of Mathematical Statistics*, 24 (1953), 254–64.

[9] Epstein, B., "Statistical problems in life testing," *Proceedings of the Seventh Annual Convention of the American Society for Quality Control*, May 1953, 385–98.

[10] Epstein, Benjamin, and Sobel, Milton, "Life testing," *Journal of the American Statistical Association*, 48 (1953), 486–502.

[11] Epstein, Benjamin, and Tsao, Chia Kuei, "Some tests based on ordered observations from two exponential populations," *Annals of Mathematical Statistics*, 24 (1953), 458–66.

[12] Feller, William, *An Introduction to Probability Theory and Its Applications*, Vol. I, New York, John Wiley and Sons, Inc. (1950).

[13] Garwood, F., "Fiducial limits for the Poisson distribution," *Biometrika*, 28 (1936), 437–42.

[14] Girshick, M. A., "Contributions to the theory of sequential analysis I," *Annals of Mathematical Statistics*, Vol. 17 (1946), 123–43.

[15] Girshick, M. A., Rubin, H., and Sitgreaves, R., "Estimates of bounded variation in particle counting," Report at September 1952 Meetings of the Institute of Mathematical Statistics.

[16] Hartley, H. O., "Testing the homogeneity of a set of variances," *Biometrika*, 31 (1940), 249–55.

[17] Kitigawa, Tosio, *Tables of the Poisson Distribution*, Tokyo, Japan, Baifukan (1951).

[18] Maguire, B. A., Pearson, E. S., and Wynn, A. H. A., "The time intervals between industrial accidents," *Biometrika*, 39 (1952), 168–80.

[19] Przyborowski, J., and Wilenski, H., "Homogeneity of results in testing samples from Poisson series," *Biometrika*, 31 (1940), 313–23.

[20] Ricker, W. E., "The concept of confidence or fiducial limits applied to the Poisson frequency," *Journal of the American Statistical Association*, 32 (1937), 349–56.

[21] Sukhatme, P. V., "On the analysis of k samples from exponential populations with special reference to the problem of random intervals," *Statistical Research Memoirs*, Vol. I (1936), 94–112.

[22] Tocher, K. D., "Extension of the Neyman-Pearson theory of tests to the discontinuous case," *Biometrika*, 37 (1950), 130–44.

[23] Wald, Abraham, *Sequential Analysis*, New York, John Wiley and Sons, 1947.

[24] Wald, Abraham, "Sequential method of sampling for deciding between two courses of action," *Journal of the American Statistical Association*, 40 (1945), 277–306.

APPLICATIONS OF THE CIRCULAR NORMAL DISTRIBUTION

E. J. GUMBEL
*Columbia University**

IN A previous paper [8] the circular normal distribution was introduced. Now its practical use will be shown. The different procedures explained in the next paragraph will be applied to typical observations taken from geophysical, vital and economic statistics. Finally, a discussion of unsolved problems will be given.

1. PROCEDURE

For simplicity's sake, we consider in the following only circular distributions over the year. The observations are grouped or can be grouped into twelve monthly periods. Then the time series is concentrated in twelve frequencies or relative frequencies p_v ($v=1, 2 \cdots 12$). In most cases p_v stands for the mean frequency of the vth month, the mean being taken for all observed years. Since the parameters of the circular normal distribution are invariant under a multiplication of all frequencies, it is irrelevant whether the absolute or relative frequencies are used. This important property allows the application of the circular normal distribution to time series which by themselves do not constitute distributions in the usual statistical sense. Let p_v be any non negative numbers of arbitrary nature or dimension, corresponding to the months; then the series $p_v / \sum_1^{12} p_v$ may be regarded as a distribution provided that the summation makes sense and may be analyzed as a circular variate.

However, in our illogical calendar system the lengths of the months vary from 28 to 31 days. If the mode occurs in July, February shows a minimum leading to an apparent asymmetry. Conversely, when the mode occurs in February, its deficiency of two days may create a hole. Since these artificial influences must be eliminated, let the observed frequencies for January, March, May, July, August, October and December be multiplied by $30/31=0.96774$ and for February by $30/28=1.07143$; let p_v' be the frequencies so adjusted. The year is thus reduced to 360 days. Therefore, the sum of the frequencies is no longer $n=\sum_{v=1}^{12} p_v$.

* Work done in part as Consultant to Stanford University and in part under grant from the Higgins Foundation.

To conserve this observed sum the adjusted frequencies are multiplied by the quotient

$$Q = \frac{p_1 + p_3 + p_5 + p_6 + p_7 + p_8 + p_{10} + p_{12}}{p_1' + p_2' + p_3' + p_5' + p_7' + p_8' + p_{10}' + p_{12}'}.$$

In this procedure the observed frequencies for April, June, September, and November are preserved while the frequencies for the other months are twice adjusted. The adjustments must be used if the differences between the frequencies are small, say of the order 10%, since they are then strongly affected by the different lengths of the months. The adjustments are not relevant if the differences between different months are large, say if the largest monthly frequency is 10 times the smallest one. The second adjustment is unnecessary if the quotient Q differs from unity by less than 1%.

The adjusted frequencies are attributed to the 15th of each month and may be traced as a linear histogram or on polar paper consisting of concentric equidistant circles and radii for each degree. The maximum frequency is traced at north and the following months are traced clockwise. After the choice of a unit distance the square roots of the twice adjusted frequencies, $\sqrt{p_\nu''}$, are plotted instead of the adjusted frequencies p_ν'' themselves. This procedure equalizes the areas of the observed and the theoretical distributions. Distributions differing with respect to the sample size n are thus traced in different scales. However, a *uniform scale* is obtained if all adjusted frequencies $\bar{p}''$ are divided by their mean $\bar{p}$. This method has the advantage that distributions with different values of n may be traced on the same scale. It has the disadvantage that the frequencies are no longer visible.

In the use of circular normal distribution

$$\phi(\alpha) = \frac{e^{k\cos(\alpha-\alpha_0)}}{2\pi I_0(k)}, \tag{1.1}$$

the mode α_0 is estimated from

$$\tan \alpha_0 = \sum_1^{12} \sin \alpha_\nu \Big/ \sum_1^{12} \cos \alpha_\nu \tag{1.2}$$

where

$$\sum_1^{12} \cos \alpha_\nu = \text{July} - \text{Jan.} + 0.86603(\text{Aug.} - \text{Feb.} + \text{June} - \text{Dec.}) \tag{1.3}$$
$$+ 0.5(\text{Sept.} - \text{Mar.} + \text{May} - \text{Nov.}),$$

$$\sum_{1}^{12} \sin \alpha_\nu = \text{Oct.} - \text{Apr.} + 0.86603(\text{Sept.} - \text{Mar.} + \text{Nov.} - \text{May}) \tag{1.4}$$
$$+ 0.5(\text{Aug.} - \text{Feb.} + \text{Dec.} - \text{June}).$$

and the months are written instead of their frequencies p_ν''. This determines α_0 only up to 180°. The exact location is found from the conventional diagram for the signs of the trigonometric functions. In the estimation of α_0 it is sufficient to calculate to whole degrees of the angle.

The parameter k is estimated from

$$\bar{a} = \frac{1}{n}\sqrt{\left(\sum_{1}^{12} \sin \alpha_\nu\right)^2 + \left(\sum_{1}^{12} \cos \alpha_\nu\right)^2} \tag{1.5}$$

with the help of Table II[1] which gives k as function $\bar{a}$. The parameters for the reduced values $p_\nu''/\bar{p}$ are obtained in an analogous manner. The adjustments for months of equal lengths may be introduced into (1.3) and (1.4). This leads to

$$\sum_{1}^{12} \cos \alpha_\nu = 0.96677(\text{Jul.} - \text{Jan.}) + 0.86603(\text{June}) \tag{1.6}$$
$$+ 0.83726(\text{Aug.} - \text{Dec.}) + 0.5(\text{Sept.} - \text{Nov.})$$
$$+ 0.48339(\text{May} - \text{Mar.}) - 0.92789 \text{ Feb.},$$

$$\sum_{1}^{12} \sin \alpha_\nu = 0.96677 \text{ Oct.} + 0.86603(\text{Nov.} + \text{Sept.}) \tag{1.7}$$
$$+ 0.48339(\text{Aug.} + \text{Dec.}) - \text{Apr.}$$
$$- 0.83726(\text{Mar.} + \text{May}) - 0.53571 \text{ Feb.} - 0.5 \text{ June}.$$

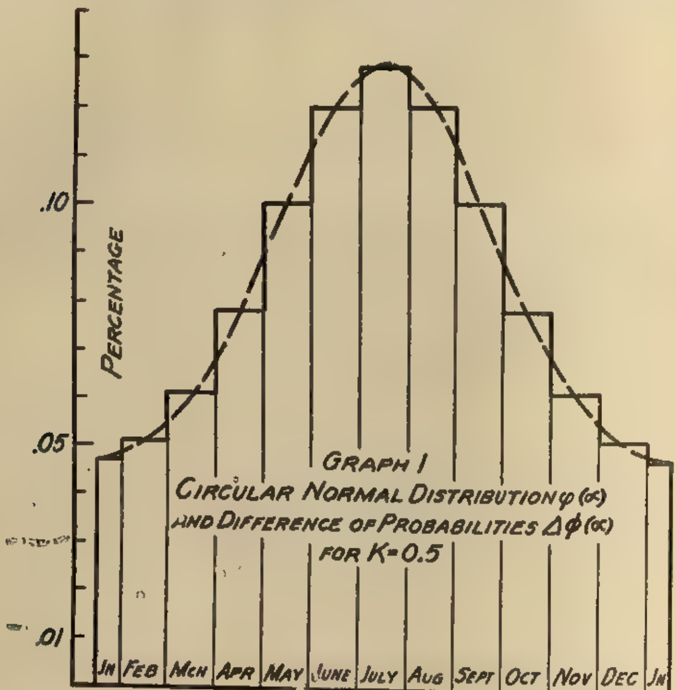
Since the parameters α_0 and k are invariant under a multiplication of the frequencies no adjustment is then necessary to conserve the sum of the observations. However, for the comparison to the theory the observations must again be adjusted for months of equal lengths and for the conservation of the observed sum. Therefore, this system is not used in the examples that follow.

In another analytic procedure a constant value equal to the minimum monthly value $\breve{p}$ is assumed to hold throughout the year. This component is treated as a circular uniform distribution. The other component consisting of the observed frequencies $p_\nu - \breve{p}$ is considered as a circular normal distribution with $n' = n - 12\breve{p}$. However, up to

[1] Tables with Roman numbers refer to the previous publication [8].

now no case has occurred in which this procedure gives a better fit than the previous one.

The theoretical values corresponding to the observations $\sqrt{p_\nu}$ are obtained after multiplying the values given in Table III by 0.28868 $\sqrt{n}$. If k is nearly equal to one of the values given in Table II, it is sufficient to use a slide rule. The theoretical values corresponding to reduced frequencies $p_\nu/\bar{p}$ are taken from III without any multiplication.



The theoretical values are plotted on the polar diagram as points. A continuous symmetrical curve is obtained by joining the points to the left and right of the mode by the same parts of a French ruler. If the modal direction is practically zero or a multiple of 30 degrees, i.e. if the observed mode is either conserved or shifted by these amounts, we may use differences of probabilities $\Delta\Phi(\alpha)$ instead of the densities $\phi(\alpha)$. This procedure leads to a comparison of the observed with the theoretical wedge diagram. It has the advantage that the conventional criteria for the goodness of fit can be used.

The differences of probabilities (Table 1) were obtained by taking

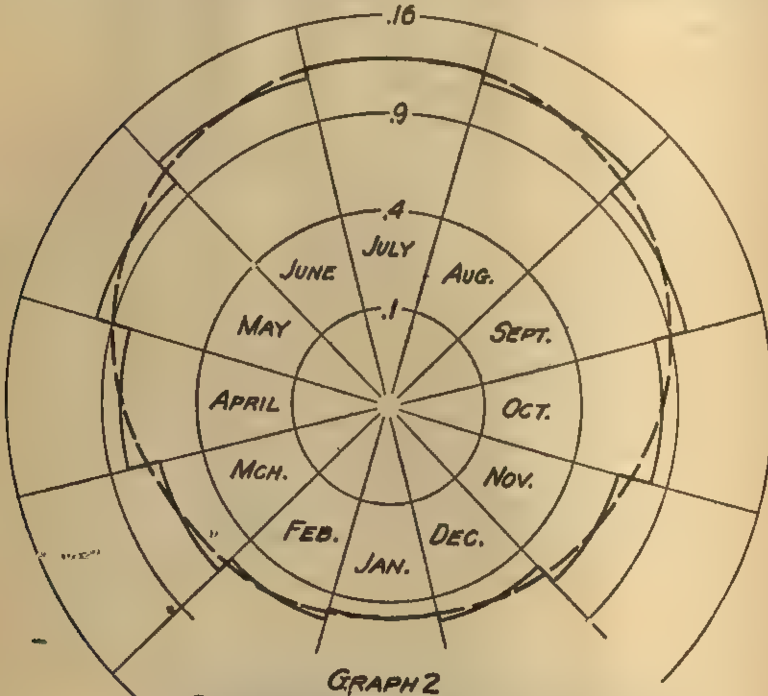
TABLE 1

THE CIRCULAR NORMAL PROBABILITY DIFFERENCES $\Delta\Phi(\alpha)$ FOR 12 AREAS, OF 30°, CENTERED ON:

k	Mean	±30°	±60°	±90°	±120°	±150°	±180°
0.0	.08333	.08333	.08333	.08333	.08333	.08333	.08333
0.1	.09176	.09056	.08735	.08314	.07912	.07630	.07530
0.2	.10054	.09793	.09111	.08255	.07476	.06953	.06770
0.3	.10962	.10539	.09458	.08158	.07031	.06304	.06058
0.4	.11895	.11286	.09774	.08024	.06581	.05690	.05394
0.5	.12846	.12032	.10054	.07858	.06133	.05110	.04780
0.6	.13810	.12768	.10296	.07662	.05690	.04570	.04217
0.7	.14782	.13491	.10501	.07440	.05256	.04069	.03704
0.8	.15755	.14196	.10666	.07196	.04837	.03608	.03239
0.9	.16726	.14880	.10793	.06933	.04434	.03186	.02822
1.0	.17690	.15539	.10882	.06656	.04049	.02804	.02449
1.1	.18644	.16171	.10934	.06369	.03686	.02459	.02118
1.2	.19584	.16772	.10953	.06076	.03345	.02149	.01826
1.3	.20507	.17345	.10938	.05781	.03026	.01872	.01569
1.4	.21413	.17885	.10895	.05484	.02730	.01627	.01344
1.5	.22298	.18395	.10824	.05190	.02457	.01410	.01149
1.6	.23163	.18874	.10729	.04901	.02206	.01219	.00979
1.7	.24007	.19320	.10613	.04617	.01978	.01052	.00833
1.8	.24829	.19738	.10476	.04344	.01768	.00906	.00707
1.9	.25630	.20127	.10324	.04078	.01578	.00778	.00599
2.0	.26409	.20487	.10158	.03822	.01407	.00668	.00506
2.1	.27167	.20822	.09979	.03578	.01252	.00572	.00427
2.2	.27905	.21130	.09790	.03345	.01113	.00489	.00360
2.3	.28623	.21415	.09592	.03124	.00988	.00418	.00303
2.4	.29322	.21677	.09388	.02914	.00876	.00353	.00255
2.5	.30003	.21917	.09179	.02718	.00776	.00304	.00214
2.6	.30666	.22138	.08966	.02528	.00687	.00258	.00179
2.7	.31312	.22338	.08752	.02352	.00607	.00220	.00150
2.8	.31942	.22522	.08535	.02186	.00536	.00187	.00126
2.9	.32557	.22688	.08318	.02031	.00474	.00158	.00105
3.0	.33157	.22838	.08100	.01886	.00418	.00135	.00088
3.1	.33744	.22974	.07884	.01750	.00369	.00114	.00073
3.2	.34317	.23096	.07670	.01624	.00325	.00096	.00061
3.3	.34878	.23204	.07458	.01505	.00286	.00082	.00051
3.4	.35427	.23301	.07249	.01394	.00252	.00069	.00043
3.5	.35964	.23386	.07042	.01292	.00222	.00058	.00036
3.6	.36490	.23460	.06839	.01196	.00196	.00049	.00030
3.7	.37006	.23524	.06639	.01107	.00172	.00042	.00025
3.8	.37513	.23579	.06442	.01026	.00151	.00035	.00020
3.9	.38009	.23625	.06251	.00949	.00133	.00030	.00017
4.0	.38497	.23662	.06063	.00877	.00117	.00025	.00014
4.1	.38976	.23692	.05879	.00812	.00102	.00021	.00012
4.2	.39446	.23714	.05700	.00750	.00090	.00018	.00010

half of the successive differences of the probability function calculated to 9 decimals and attributing them to the midpoints of the intervals.

Graphs 1 and 2 show a wedge diagram obtained from Table 1 and the continuous distribution traced on linear and on aequiareal polar scales. Since the observations can only be traced in the wedge form, while the theoretical distribution can also be traced as a continuous curve, it is



GRAPH 2
CIRCULAR NORMAL DISTRIBUTION $\phi(\alpha)$
AND DIFFERENCE OF PROBABILITIES $\Delta\phi(\alpha)$
TRACED ON AN EQUIVALENT POLAR SCALE
K=0.5

desirable to compare the appearance of one and the same distribution in the two forms. The graphs also show the type of deviations that are inevitable if an observed wedge diagram is compared to a continuous distribution. With the exception of the modal and anti-modal months, the curves representing the theoretical distribution intersect the observed wedge diagram near the middle of each month and the wedge diagram falls short (exceeds) the theoretical curve at the beginning (end) of each month. The tracing of $\Delta\Phi(\alpha)$ is a graphical alternative to the tracing of the distribution $\phi(\alpha)$. In general it has the disad-

vantage that the "months" considered therein do not coincide with the months in the calendar.

The different analytical and graphical procedures outlined will now be applied to numerical data.

2. GEOPHYSICAL OBSERVATIONS

The following examples deal with rainfall, run-off and evaporation, i.e. the hydrologic cycle and temperatures. The *amounts of rainfall per month* form a genuine distribution. For the Esopus watershed which is essential to the water supply of New York City, Table 2 shows the mean monthly rainfall. The modal (anti-modal) month is written in the first (last) line. The second column contains the observed numbers in inches, the third introduces the adjustments for length of months. The sum of the adjusted frequencies is 30.665″ while the observed sum for the same months before adjustment is 31.36″. Consequently the adjusted frequencies are multiplied by $31.36/30.665 = 1.02263$, resulting in the fourth column which conserves the observed total mean annual rainfall $n = 48.91''$. The square roots of the numbers in the fourth column starting with July are plotted in the wedge diagram, Graph 3. The variations indicate a systematic circular behavior starting with a mode in July and decreasing to the anti-mode in January with the exception of the rainfalls in November which exceed those in October.

TABLE 2

MEAN MONTHLY RAINFALL IN INCHES: ESOPUS WATERSHED

1 Month	2 Obs.	3 First Adj.	4 Sec. Adj.	1 Month	2 Obs.	3 First Adj.	4 Sec. Adj.
July	4.69	4.539	4.64	—	—	—	—
June	4.54	—	4.54	Aug.	4.58	4.432	4.53
May	4.31	4.171	4.27	Sept.	4.50	—	4.50
Apr.	4.23	—	4.23	Oct.	4.03	3.900	3.99
Mar.	3.85	3.726	3.81	Nov.	4.28	—	4.28
Feb.	3.05	3.268	3.34	Dec.	3.45	3.339	3.42
—	—	—	—	Jan.	3.40	3.290	3.36

To obtain the theoretical distribution we first calculate the parameters. Formulas (1.3) and (1.4) lead from Table 2, column 4 to the sums

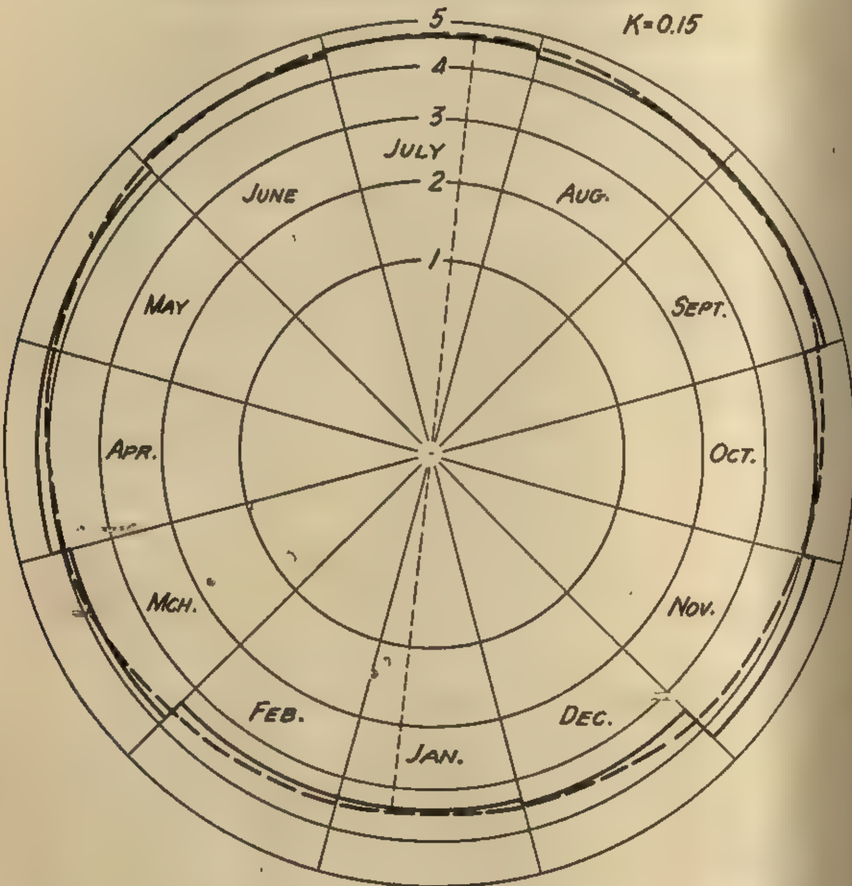
$$\sum_{1}^{12} \cos \alpha_\nu = 3.62553; \qquad \sum_{1}^{12} \sin \alpha_\nu = 0.40142.$$

Consequently

$$\tan \alpha_0 = 0.11067; \qquad \alpha_0 = 6^0.$$

Since α is counted from 15 July, this means a shift of the mode to 21 July. It is due to the fact that the frequencies for the latter part of the

GRAPH 3
MONTHLY MEAN RAINFALL ESOPUS WATERSHED



year show a slight predominance over the first part (20.72″ against 20.19″). Formula (1.5) leads to

$$\bar{a} = 3.64766/48.91 = 0.07458.$$

Table II gives the values

$$\bar{a} = 0.07, \quad k = 0.1403; \qquad \bar{a} = 0.08, \quad k = 0.1605.$$

Linear interpolation leads to $k=0.1495$. It is sufficiently accurate to choose $k=0.15$. This small value of the parameter is due to the fact that the differences between the frequencies are quite small.

From Table III we obtain radii vectors $\psi(\alpha)\sqrt{12/n}$ by interpolating between $k=0.1$ and $k=0.2$. These values are given in Table 3, column 2. Multiplication by $\sqrt{48.91/12}=2.0189$ leads to the radii vectores (column 3), which are plotted on the original grid of the polar paper to the left and to the right of the modal value at 21 July. The fit of the theory to the observations leaves nothing to be desired.

TABLE 3

CALCULATION OF RADII VECTORS

1	2	3
Angle	$\psi(\alpha)\sqrt{12/n}$	$\psi(\alpha)$
0°	1.075	2.170
20°	1.070	2.160
40°	1.051	2.122
60°	1.035	2.090
80°	1.012	2.043
100°	0.9840	1.987
120°	0.9603	1.939
140°	0.9414	1.901
160°	0.9293	1.876
180°	0.9252	1.868

The *numbers of occurrences of rainfall* of 1″ or more per hour also form a genuine distribution. The observed numbers taken from Dyck [7] for 156 stations in the United States, 1908–37, are given at the bottom of the linear Graph 4. The sum of the adjusted frequencies is 4366.5 while the corresponding sum without adjustment is 4502. Multiplication of the adjusted numbers by 1.03103 conserves the observed total sum 7235. The twice adjusted numbers are given at the top of Graph 4. The adjustments are less than 1% except for February where the observed value is decreased by more than 10%. The mode in July is more than 14 times the minimum in January. Therefore, the parameter k is relatively high, $k=1.34$. The calculation of the parameters in this and all following examples is given in Table 9. The distribution is slightly skewed to the right; consequently, the mode is shifted from 15 to 18 July.

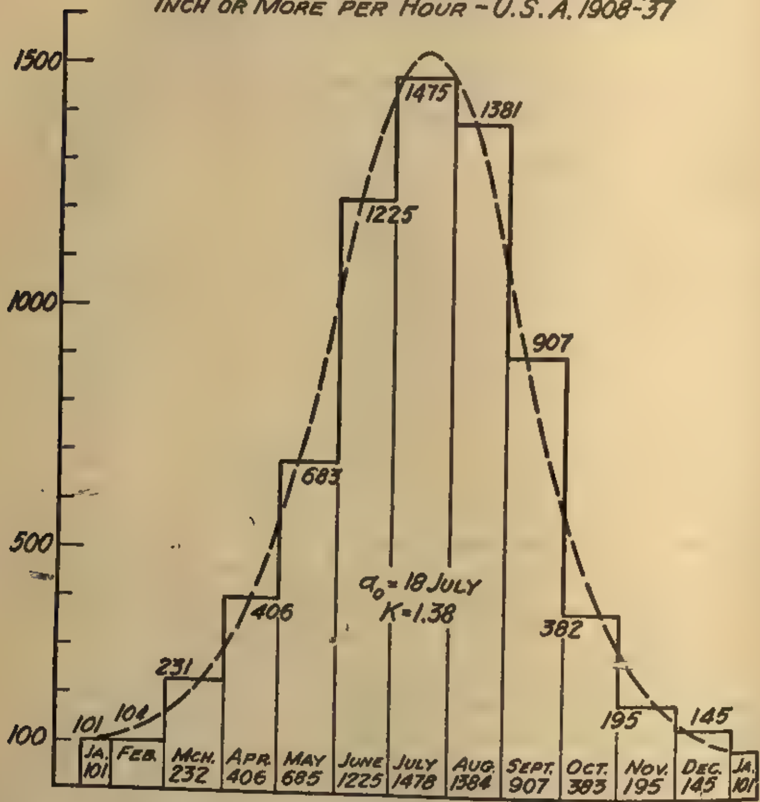
The theoretical distribution is obtained by the same procedure as before. We interpolate $\psi(\alpha)\sqrt{12/n}$ from Table III, multiply these volumes by $\sqrt{n/12}=24.554$ to obtain $\psi(\alpha)$. The squares of these num-

bers are plotted in Graph 4 and traced in a continuous curve.

The *monthly runoff* in inches for the watershed of the Derwent River at Yorkshire Bridge, Derbyshire (England) taken from [14], p. 397, is given in Table 4 for the 43 years June 1905 to May 1948. The monthly percentages of the annual runoff form a distribution. Therefore the monthly runoff can be analyzed as a cyclical variate.

GRAPH 4

NUMBER OF OCCURANCES OF RAINFALL ONE INCH OR MORE PER HOUR—U.S.A. 1908–'37



The second column gives the runoff for the 43 years. The third column gives the monthly means adjusted for length of months, the fourth conserves the observed yearly sum $n=36.28''$. The distribution is slightly asymmetrical since the mode is in January while the anti-mode is in June. The observations and the theory are traced in equivalent polar and in linear scales in Graph 5. The fit is quite satisfactory, since the deviations between theory and observations do not show any

systematic trend. The authors of the article [14] assumed that the differences between the months are due to chance. This opinion implies a uniform distribution of the monthly runoff which seems highly improbable, to say the least.

The circular normal distribution cannot be applied to rivers with two regimes, one derived from rainfalls, say in Autumn, and the other from the melting of snow, say in Spring. These conditions lead to a bi-modal distribution.

TABLE 4

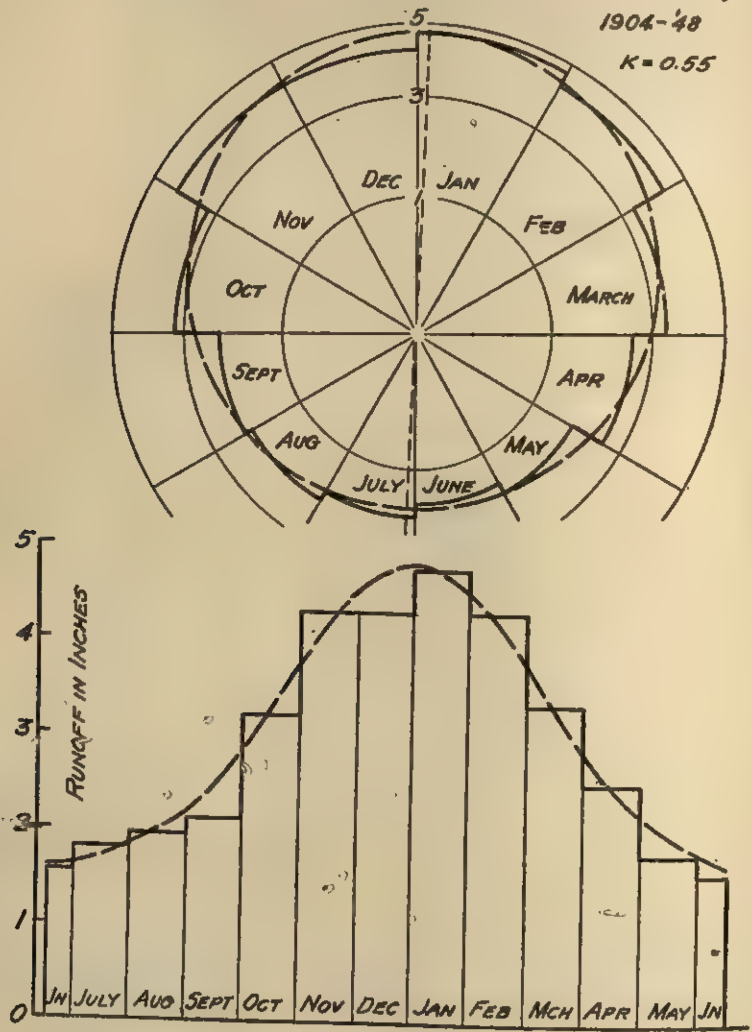
MONTHLY RUNOFF IN INCHES: DERWENT RIVER 1905–48

1	2	3	4	1	2	3	4
Month	Sum	Runoff	per Month	Month	Sum	Runoff	per Month
		First Adj.	Sec. Adj.			First Adj.	Sec. Adj.
Jan.	210.29	4.733	4.81	Dec.	191.47	4.309	4.38
Feb.	171.87	4.282	4.35	Nov.	185.17	4.306	4.31
Mar.	147.85	3.327	3.38	Oct.	142.28	3.202	3.26
Apr.	106.84	2.485	2.49	Sept.	91.57	2.130	2.13
May	78.85	1.775	1.80	Aug.	86.54	1.948	1.98
June	67.27	1.564	1.56	July	80.09	1.802	1.83

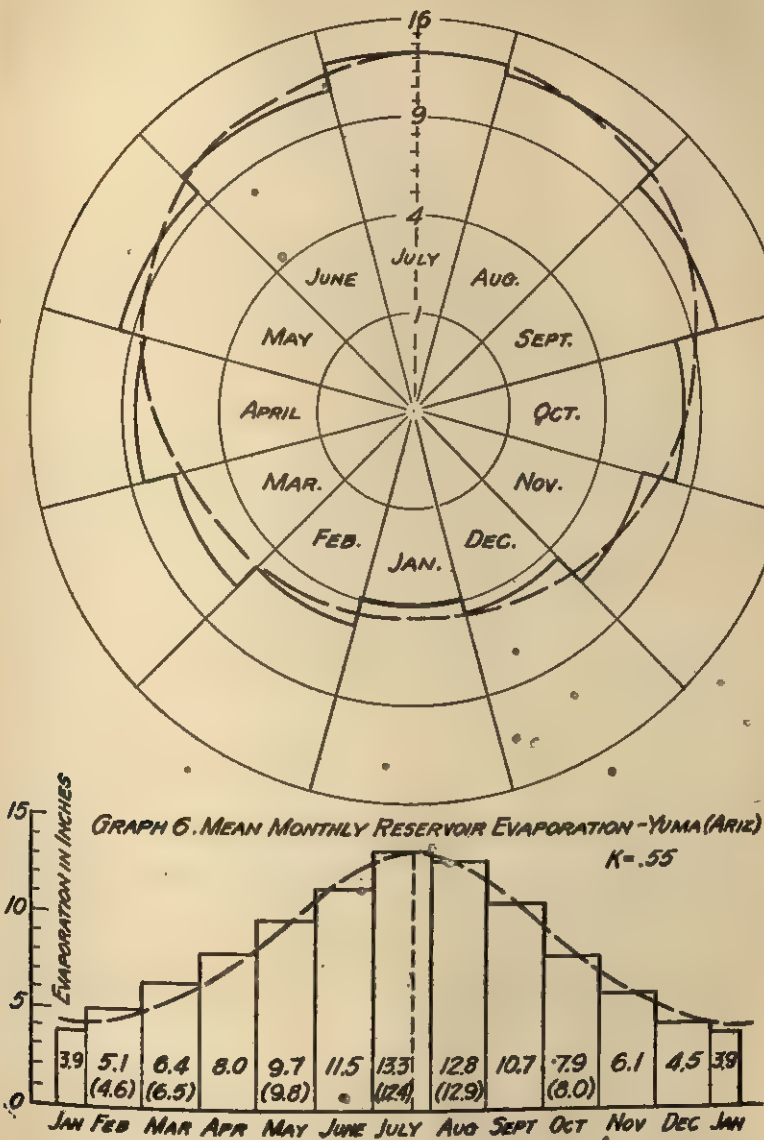
The *monthly evaporations* from a reservoir do not form a distribution but they may be analyzed as the preceding examples for the reasons given above. Since the evaporations depend upon temperature, the maximum occurs in summer and the minimum in winter. The observations in inches for Yuma, Arizona, taken from the Hydrology Handbook [9] p. 127, and the twice adjusted data are given in the linear Graph 6. The mode occurs in July and the anti-mode in January. However, the evaporations in the second half of the year are slightly stronger than in the first one. Consequently there is a slight shift in the mode to 21 July. The theory leads to a very good fit to the observations.

The *mean monthly temperatures* at a given place do not constitute a distribution, since temperatures cannot be added. However, it is certainly legitimate to trace them as a function of time. This is done in Graph 7 for the mean monthly temperatures in Boston (Mass.) taken from Conrad [2] p. 110. A reduction for months of equal lengths does not seem to be necessary. The circular distribution, used to reproduce the data, gives an excellent picture.

GRAPH 5 MONTHLY RUNOFF - DERVENT RIVER (ENGLAND)



We conclude from the previous examples, which are illustrative of a wide range of problems, that the circular normal distribution is an efficient tool for the analysis of meteorological phenomena, even in cases which are not actually distributions.

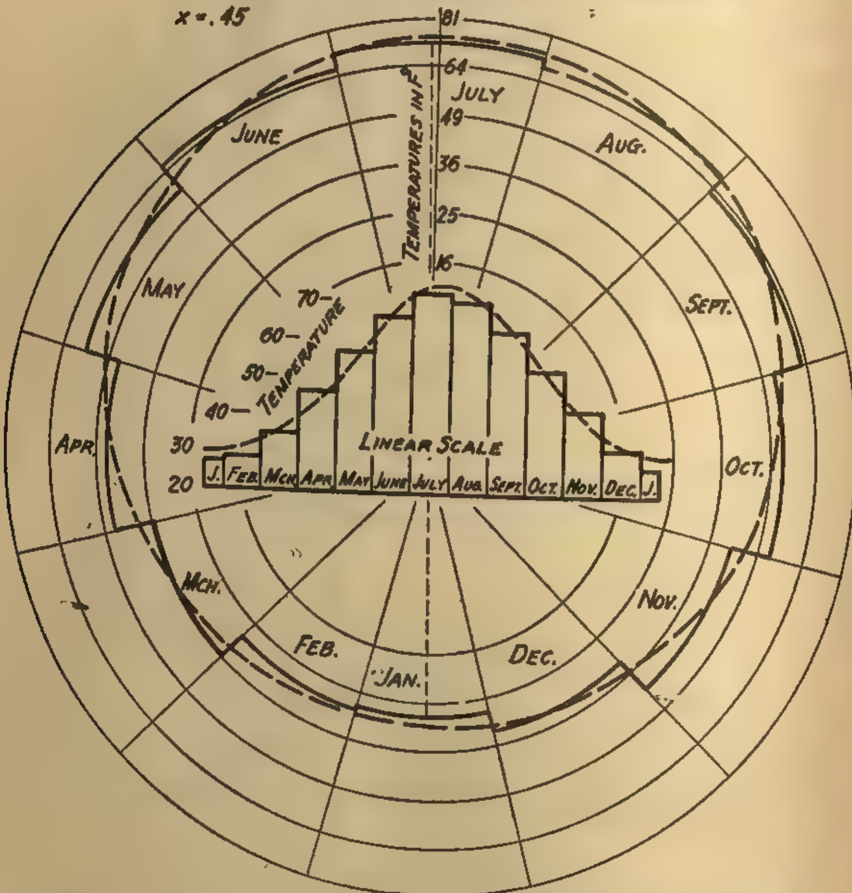


GRAPH 6. MEAN MONTHLY RESERVOIR EVAPORATION - YUMA (ARIZ)

3. APPLICATIONS TO VITAL STATISTICS

The percentage of persons who die each month forms a distribution. Since the population is growing in most countries it is customary in vital statistics to present instead the death rates, i.e. the number of deaths divided by the time lived through by the respective popula-

GRAPH 7
MONTHLY TEMPERATURES - BOSTON (Conrad p.110)



tion. These *monthly death rates* per 1000 have the dimension of the reciprocal of the time and can be compared month by month, and year by year. They may be treated for our purposes as if they constituted a distribution, since the death rates are proportional to the number of deaths for a stationary population.

The monthly means of the observed and adjusted death rates in the United States, September 1946 to August 1951 are given in Table 5, where the mode and anti-mode are given in the first and last lines respectively.

TABLE 5

DEATH RATES U. S. A. SEPT. 1946–AUG. 1951

1 Month	2 Obs.	3 First Adj.	4 Sec. Adj.	1 Month	2 Obs.	3 First Adj.	4 Sec. Adj.	5 Theory
Feb.	10.66	11.42	11.64	—	—	—	—	10.82
Jan.	10.50	10.16	10.35	Mar.	10.80	10.45	10.65	10.68
Dec.	10.42	10.08	10.27	Apr.	10.24	—	10.24	10.31
Nov.	9.70	—	9.70	May	9.60	9.29	9.46	9.81
Oct.	9.42	9.12	9.29	June	9.42	—	9.42	9.34
Sept.	9.08	—	9.08	July	9.18	8.88	9.05	9.00
—	—	—	—	Aug.	8.96	8.67	8.88	8.88
						$n=$	117.98	117.98

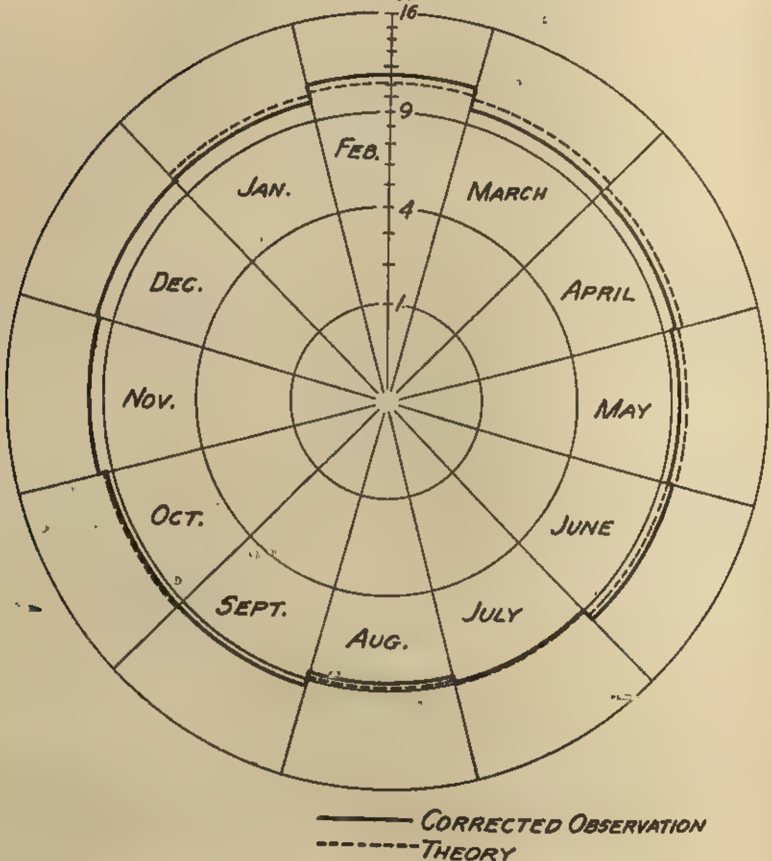
The two adjustments leave the sum at $n=117.98$, shift the mode from March to February and leave the anti-mode in August. Thereby the distribution is made more symmetrical. The square roots of the corrected rates are traced on the polar Graph 8.

It is sufficiently accurate to use $k=0.1$. This value is rather small because the monthly variations are also small. Since both trigonometric sums are negative, the mode obtained from $\tan \alpha_0 = 0.58487$ is $\alpha_0 = 210°$. The mode on 15 February is identical with the observed mode. Therefore, we may use the differences of probabilities $\Delta\Phi(\alpha)$ in Table 1 to obtain the theoretical rates. The values $117.98\Delta\Phi(\alpha)$ are given in the last column of Table 5. The square roots of these values are plotted on the Graph 8. The agreement between theory and observation is quite close. No systematic deviations are visible except that the theoretical mode is slightly smaller than the observed one.

Infant death rates are a very important part of vital statistics since they reflect Social Hygiene. These death rates are proportional to the number of infant deaths for a stationary population and can therefore be analyzed by the circular theory. The mean rates for the different months twice adjusted are shown at the bottom of the linear Graph 9. The mode and the decrease of the frequencies toward the anti-mode are quite systematic: the monthly variations are very small. Consequently, the value of the parameter k also turns out to be very small and the theoretical distributions traced in Graph 9 resemble a uniform distribution.

The cyclical variation of the number of deaths over the years is a very complex result of different causes of deaths having maxima at different times. For example, influenza has a mode in January, pneumonia in February, scarlet fever and meningitis in March, measles in

GRAPH 8
DEATH RATES U.S.A. SEPT. 1946 TO AUGUST 1951
K=0.105

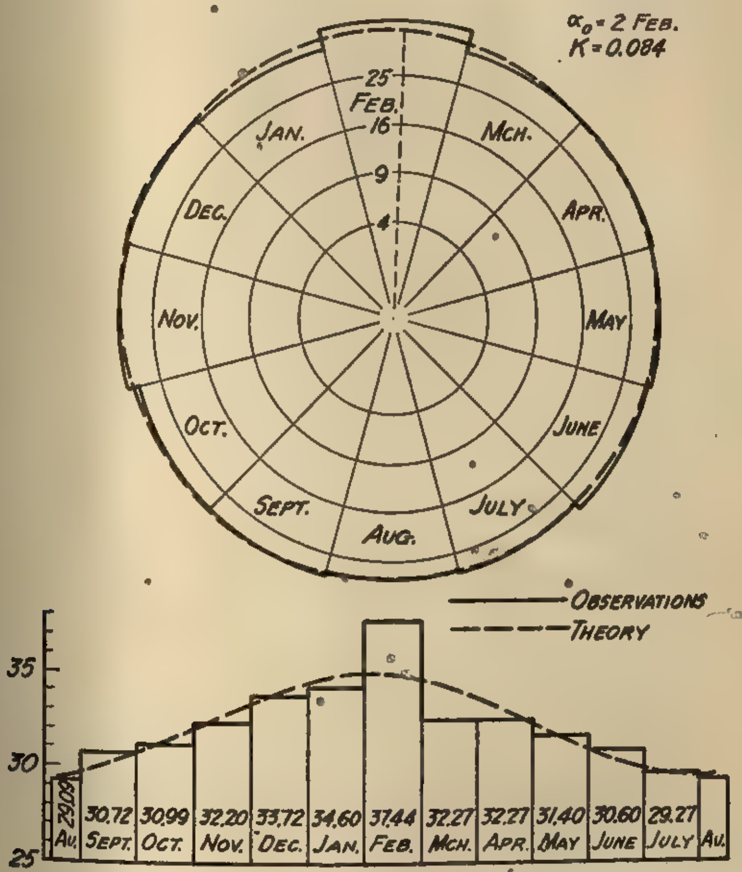


April, dysentery in July, typhoid fever in August, poliomyelitis in September, diphtheria in October, and tularemia in December. The day with maximum probability of deaths should be considered as critical. The popular idea of the existence of critical days for certain illnesses is completely legitimate.

The mean percentages of *deaths of children* under 1 year of age in the United States, August 1947 to July 1951, due to *pneumonia and influenza* are given in Table 6. These percentages are proportional to the number of deaths from these causes for a constant number of

GRAPH 9

INFANT DEATH RATES U.S.A. SEPT. 1945 TO AUG. 1951



children who died. Therefore they may be represented by the circular distribution. The sum of the observations is $n = 114.27$; the sum of the corrected data is $n' = 113.32$. Since the second adjustment would have amounted to less than 1%, it was not applied.

TABLE 6

PERCENTAGE OF DEATH OF CHILDREN DUE TO PNEUMONIA AND INFLUENZA, U. S. A. 1947–1951

1 Month	2 Frequency Observed	3 Adjusted	1 Month	2 Frequency Observed	3 Adjusted	4 Theor. Rate
Feb.	16.07	17.22	—	—	—	15.38
Jan.	14.49	14.02	Mar.	13.97	13.52	14.27
Dec.	10.31	9.98	Apr.	11.46	11.46	11.60
Nov.	9.40	9.40	May	8.35	8.08	8.74
Oct.	7.58	7.34	June	6.05	6.05	6.57
Sept.	6.02	6.02	July	5.05	4.89	5.33
—	—	—	Aug.	5.52	5.34	4.93
			Sum	114.27	113.32	113.33

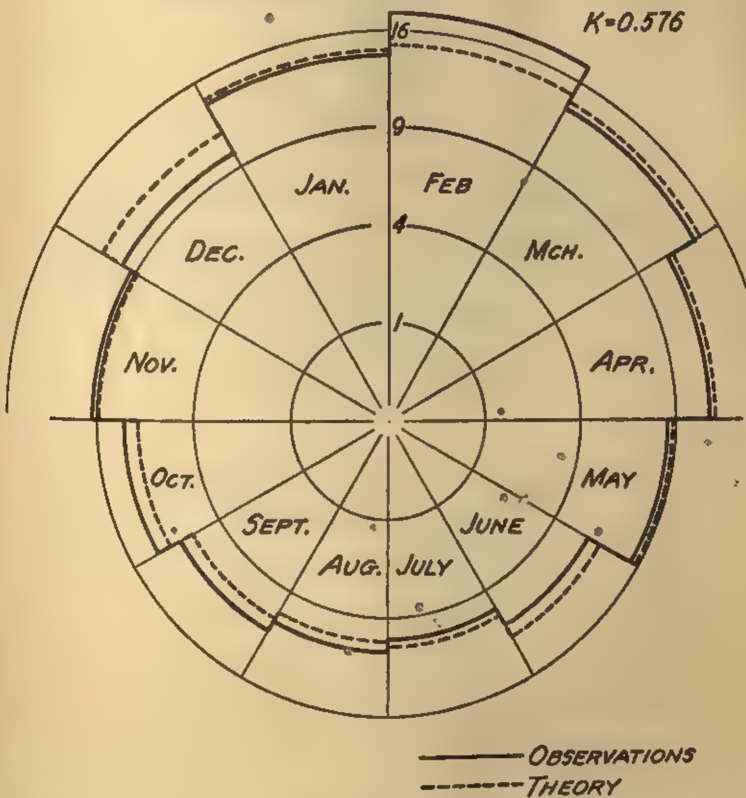
The distribution is not quite symmetrical. The mode is located in February but the anti-mode in July (instead of August). Since the mode falls on 15 February we use again the probability Table 1. The observed and theoretical wedge diagrams are compared in Graph 10. The first may be accepted although the theoretical mode falls short of the observed one.

Pearl [12], p. 153 gives the *percentage of total infant mortality due to diarrhea and enteritis* in the United States for each month, 1933–1935. Since these percentages are proportional to the monthly number of deaths from these causes for a stationary population with constant infant mortality, they may be treated as a circular distribution. The mean monthly percentages for the three years adjusted first for months of equal lengths and then for the conservation of the sum are given in the linear Graph 11. The distribution is slightly skew, since the mode is in July, whereas the anti-mode is in February. This produces a shift of the theoretical mode to the beginning of August. The fit of the theory to the observation shown in the circular Graph 11 is again very satisfactory.

The observed and twice adjusted *numbers of persons drowned in each month* in the United States in 1946 are given respectively in the bottom and in the end of the linear Graph 12. The square roots of the corrected numbers are traced in the wedge diagram. Evidently the maximum must lie somewhere in the summer. However, some unfortunate persons drown even in winter by breaking through the ice.

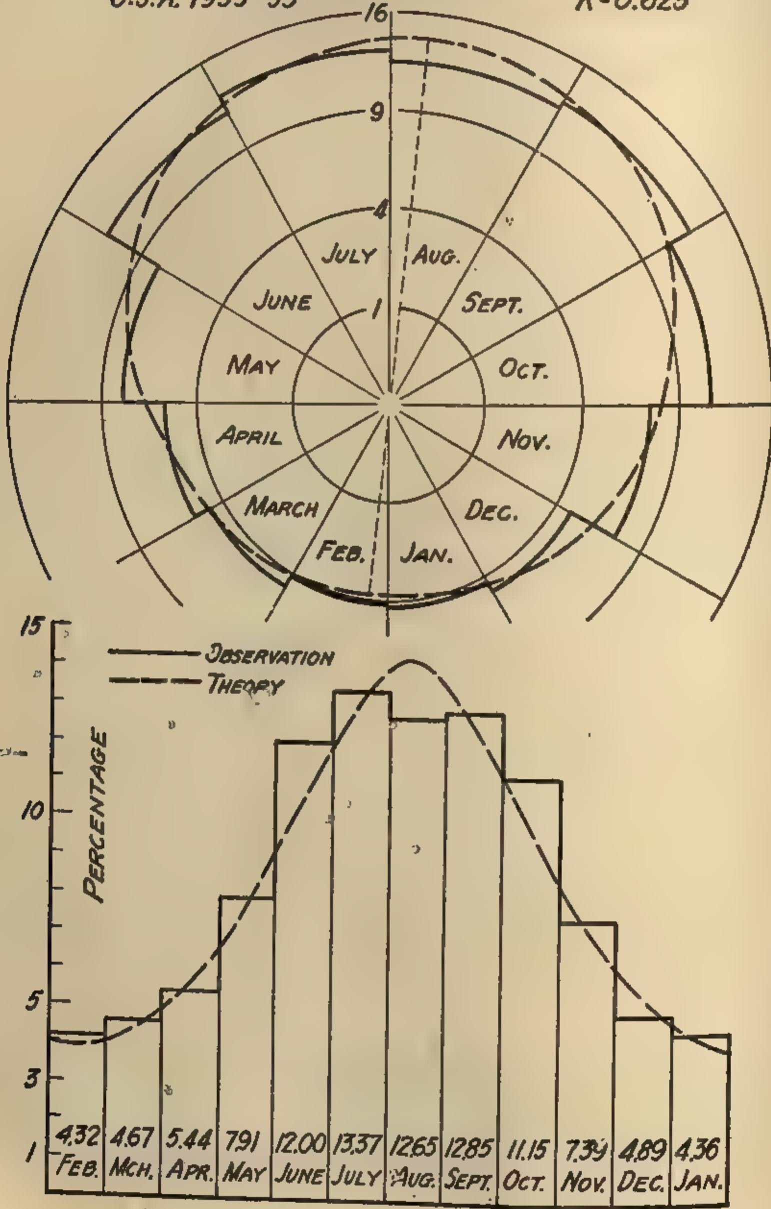
The circular character of the distribution is thus clear. However, the distribution is not quite symmetrical. The mode is in July, but the anti-mode is in December. The frequencies for the beginning of the year are stronger than for the end. Therefore, the theoretical mode is at 26 June instead of 4 July where it should be according to popular customs. The mode of the theoretical curve is smaller than the observed

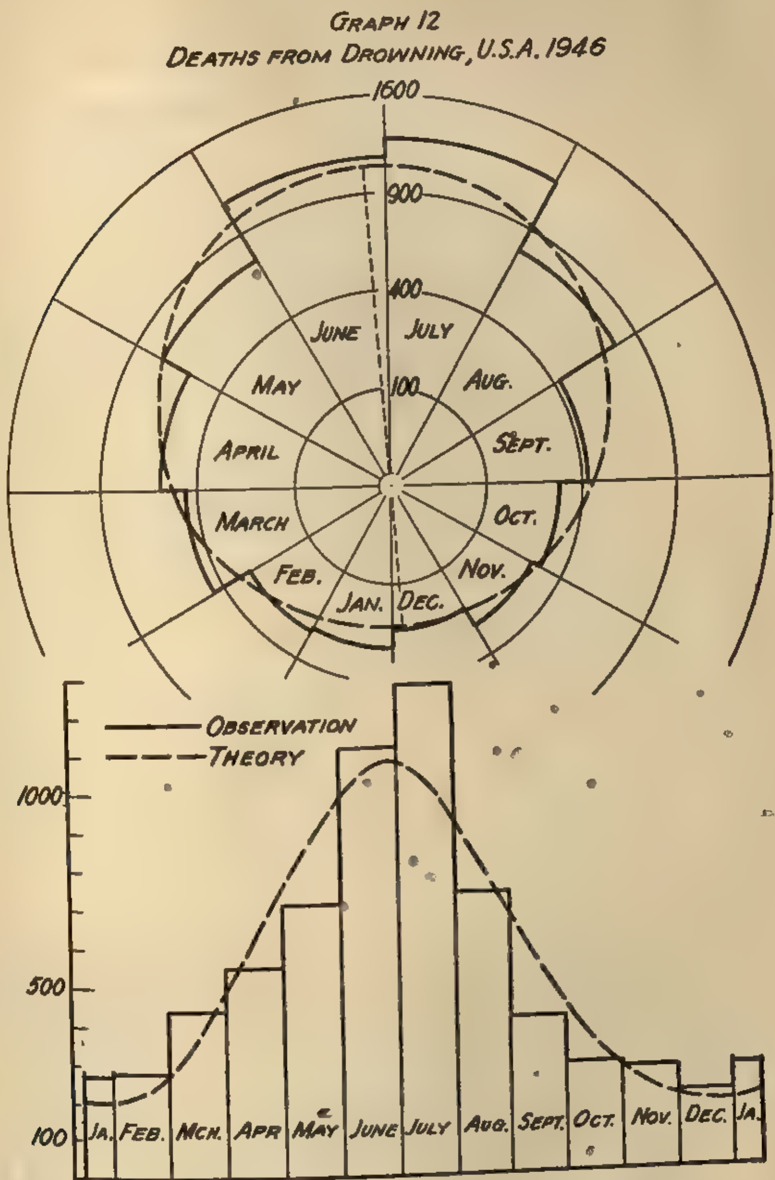
GRAPH 10
PERCENT OF PNEUMONIA AND INFLUENZA IN INFANT DEATHS U.S.A. AUG. 1947 TO JULY 1951



one. Consequently the theory gives too large values for August to October. These deviations are due to the skewness of the observed distribution caused by the heterogeneity in the data. The activities in which drowning occurred in July differ from those in December. The drowned in summer come from another population than those drowned in winter.

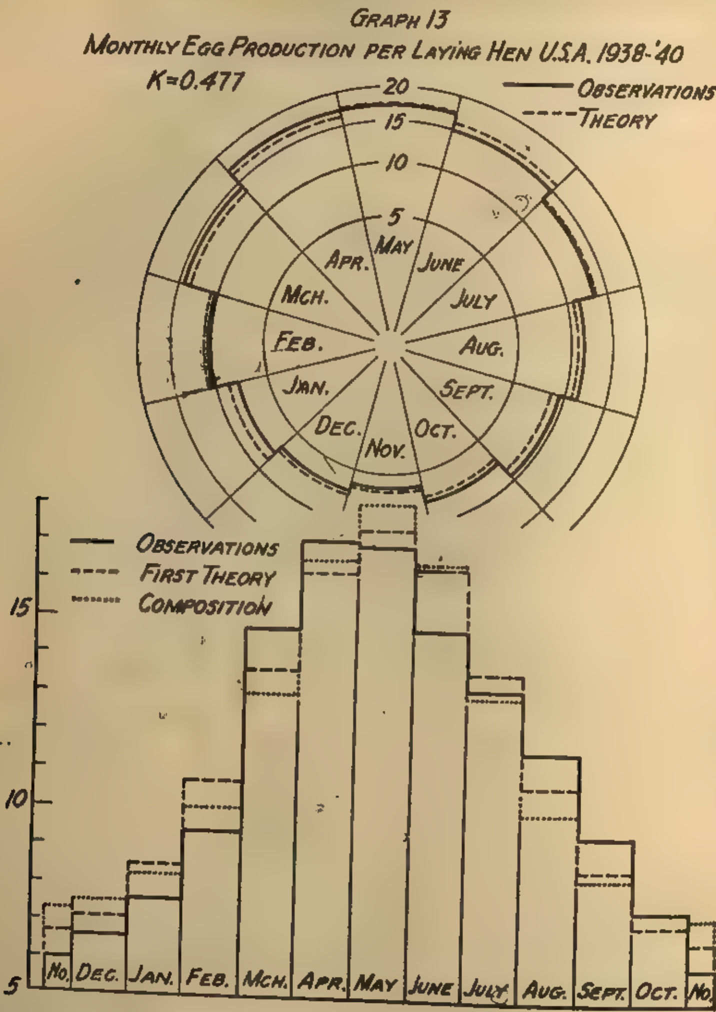
GRAPH II
PERCENT OF TOTAL INFANT MORTALITY
DUE TO DIORRHEA AND ENTERITIS
U.S.A. 1933–'35 K=0.625




GRAPH 12
DEATHS FROM DROWNING, U.S.A. 1946
1600
900
400
100
JUNE
JULY
MAY
AUG.
APRIL
SEPT.
MARCH
OCT.
FEB.
NOV.
JAN.
DEC.
OBSERVATION
THEORY
1000
500
100
JA.
FEB.
MCH.
APR
MAY
JUNE
JULY
AUG.
SEPT.
OCT.
NOV.
DEC.
JA.

4. APPLICATIONS TO ECONOMIC TIME SERIES

Other possible applications may be found in economic time series. But they entail certain difficulties which must be pointed out. First,



most series of prices, production and trade do not constitute distributions. Second, economic time series usually contain secular or cyclical movements, which the economist would want to eliminate before investigating the seasonal movements; but the process of eliminating these undesirable influences and expressing the series as a set of sea-

sonal relatives, as it is often done, leaves the series with less resemblance to a distribution than it had originally. Finally, the seasonals in many time series have configurations that cannot be adequately fitted by the assumption of a single circular distribution. These three difficulties are amplified in connection with the illustrative examples presented below.

The monthly *egg production per laying hen*, Table 7, given by Kendall [10], p. 368, is not, strictly speaking, a frequency distribution. Although the eggs produced in a period of years by a single hen, a flock of hens or even by all hens consist of discrete countable units, the process of dividing the number of eggs produced by the number of hens producing them conceals the actual egg frequencies, which are not available. However, the eggs-to-hens ratio may be regarded as a sort of distribution for an average hen or an average flock. The observed and twice adjusted data are given in Table 7 and traced in Graph 13.

TABLE 7

MONTHLY EGG PRODUCTION PER LAYING HEN U. S. A.
1938–1940

1 Month	2 Obs.	3 Adj.	1 Month	2 Obs.	3 Adj.	4 First Theory	5 Second Theory
May	17.10	16.91	—	—	—	17.20	18.13
Apr.	17.00	17.00	June	14.77	14.77	16.14	16.47
March	14.90	14.73	July	13.40	13.25	13.59	12.99
Feb.	9.53	10.43	Aug.	11.77	11.63	10.74	10.01
Jan.	7.70	7.61	Sept.	9.47	9.47	8.48	8.29
Dec.	6.67	6.66	Oct.	7.60	7.51	7.13	7.52
—	—	—	Nov.	6.03	6.03	6.69	7.30

The calculations lead to a mode on 16 May. We choose instead 15 May, a difference which is not visible in the graphs. Interpolation for $k=0.477$ in Table 1 leads to the theoretical number of eggs given in Table 7, column 4. The fit between theory and observations is good and no systematic deviations exist.

If we consider only the differences of the numbers given in Table 7, column 3 from their minimum 6.033, the sum of these numbers is $n'=63.64$. The resulting vector strength obtained from (1.5) is $\bar{a}=0.4961$. The direction of the mode remains unchanged. Table II leads to the estimate $k=1.14$. The theoretical results $6.03+63.64\Delta\Phi(\alpha)$ obtained from Table 1 are given in the last column of Table 7 and are

traced in the lower part of Graph 13. The difference in the two representations are not visible in the polar Graph 13. The distribution obtained by the composition of a uniform and a circular distribution does not show a better fit to the observations than the simple procedure used before.

The observations in the following examples are treated according to the uniform procedure outlined in paragraph 1.

The series on *automobile production* [1] is a genuine frequency distribution consisting of discrete, countable units—i.e., the number of automobiles produced in three years allocated by months. Since these are original data, no adjustments have been made for cyclical, secular or irregular influences, except those for length of months (column 3).

TABLE 8

PASSENGER CARS PRODUCTION

1 Month	2 Obs.	3 Adj.	4 Redcd.	1 Month	2 Obs.	4 Adj.	4 Redcd.
Aug.	1589	1538	1.193	—	—	—	—
Sept.	1453	1453	1.127	July	1435	1389	1.078
Oct.	1523	1674	1.144	June	1527	1527	1.185
Nov.	1251	1251	0.971	May	1197	1158	0.898
Dec.	1192	1154	0.895	Apr.	1200	1200	0.931
Jan.	1120	1084	0.841	Mar.	1223	1184	0.919
—	—	—	—	Feb.	985	1055	0.818
				Sum	15695	15667	12.000

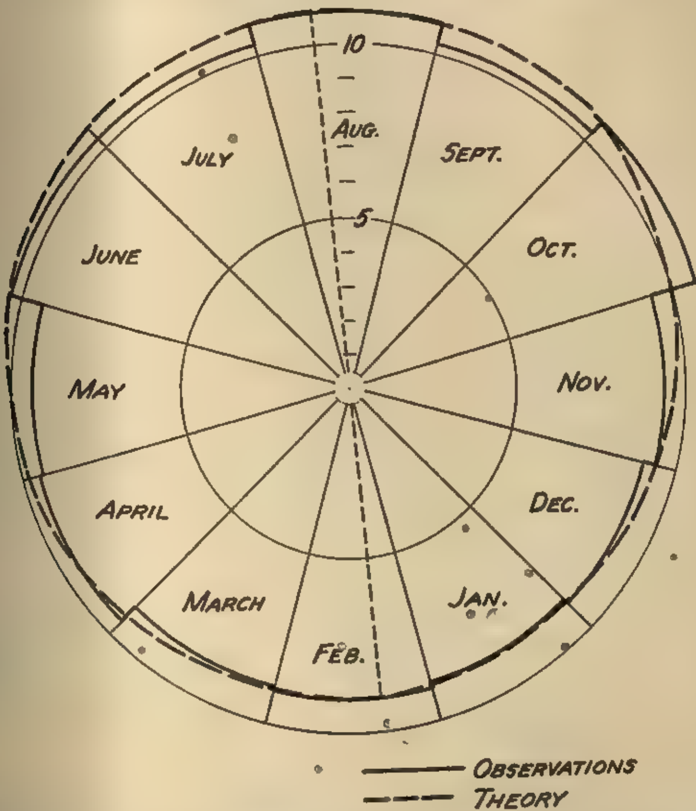
The mean monthly number of cars produced is 1308. Division of column 3 by this number leads to column 4, the sum of which is of course 12.

The irregular movements are rather pronounced. August is the modal month, but October and June exceed September and July, hence these months appear as conspicuous valleys. Since $n=12$ the values $\psi(\alpha)$ in Table III constitute the theoretical distribution. It is sufficiently accurate to use $k=0.2$. The smooth curve in Graph 14 does not give a bad fit and especially the mode and anti-mode are conserved.

The series on *pig-iron* production, taken from Croxton and Cowden [5], p. 374, consists of seasonal relatives corrected for cyclical, secular and irregular influences. The resulting wedge diagram, Graph 15 is, therefore, fairly regular. The amplitude is not very large. The antimode precedes the mode by 5 instead of by 6 months, implying a some-

what skewed distribution. The mode is well reproduced and on the whole the fit is good, since the deviations between theory and observations do not show any systematic pattern.

GRAPH 14
PASSENGER CAR PRODUCTION U.S.A. 1948-'50 IN UNIFORM SCALE



In the *sales of Misses' coats*, 1926–30 [11] p. 393, two modes and two antimodes exist, April, October and July, February, respectively. Clearly such observations cannot be fitted into the scheme of a single circular distribution.

5. OUTLOOK

The introduction of a new distribution solves certain statistical problems, but at the same time raises a considerable number of new problems, some of which are enumerated below:

TABLE 9

ESTIMATION OF PARAMETERS

1 Material	Symbol	From	Rainfall	Runoff	Evaporation	Temperature
2 Place			USA 156 Stations	Derwent	Yuma (Ar.)	Boston (Mass.)
3 Date			1908–37	1905–48		
4 Sample Size	n	Obs.	7235	36.28	99.9	595.2
5 Cosine	$\sum_1^{12} \cos \alpha_\nu$	eq. (1.3)	3997.23	− 9.355	26.081	130.19
6 Sine	$\sum_1^{12} \sin \alpha_\nu$	eq. (1.4)	237.31	2.086	2.856	21.58
7 Tangent	$\tan \alpha_0$	lines 5, 6	0.0594	− 0.223	0.1095	0.1657
8 Mode	α_0	line 7	18 July	2 Jan.	21 July	24 July
9 Sum	$n\bar{a}$	eq. (1.5)	4004.27	9.548	26.237	131.96
10 Vector Strength	$\bar{a}$	lines 4, 9	0.553	0.264	0.263	0.222
11 Parameter	k	Table II	1.338	0.548	0.545	0.455
12 Factor	$\sqrt{n/12}$	line 4	24.554	1.739	2.885	7.043
13 Table			in graph	4	in graph	—
14 Graph			4	5	6	7

1 Material	Symbol	From	Death Rate	Infant Death Rate	% Pneumonia	% Diarrhea
2 Place			USA	USA	USA	USA
3 Date			Se 46–Au 51	Se 45–Au. 51	Au. 47–Ju. 51	1933–1935
4 Sample Size	n	Obs.	117.98	383.97	113.32	101.00
5 Cosine	$\sum_1^{12} \cos \alpha_\nu$	Eq. (1.3)	− 5.385	−15.241	−27.232	26.732
6 Sine	$\sum_1^{12} \sin \alpha_\nu$	eq. (1.4)	− 3.082	− 4.545	−15.507	12.934
7 Tangent	$\tan \alpha_0$	lines 5, 6	0.572	0.298	0.569	0.484
8 Mode	α_0	line 7	15 Feb.	2 Feb.	15 Feb.	11 Aug.
9 Sum	$n\bar{a}$	eq. (1.5)	6.204	15.746	31.338	29.696
10 Vector Strength	$\bar{a}$	lines 4, 9	0.053	0.041	0.276	0.294
11 Parameter	k	Table II	0.105	0.082	0.576	0.625
12 Factor	$\sqrt{n/12}$	line 4	—	5.657	3.071	2.901
13 Table			5	in graph	6	in graph
14 Graph			8	9	10	11

1 Material	Symbol	From	Drowning	Eggs per Hen	Car Prod.	Pig Iron
2 Place			USA	USA	USA	USA
3 Date			1946	1938–40	1948–50	1936
4 Sample Size	n	Obs.	6632	136	12.00	12.00
5 Cosine	$\sum_1^{12} \cos \alpha_\nu$	eq. (1.3)	2398	16.502	1.045	0.3635
6 Sine	$\sum_1^{12} \sin \alpha_\nu$	eq. (1.4)	−850	−26.915	0.494	− 0.4756
7 Tangent	$\tan \alpha_0$	lines 5, 6	−0.354	−1.631	0.473	−1.308
8 Mode	α_0	line 7	26 June	16 May	10 Aug.	22 May
9 Sum	$n\bar{a}$	eq. (1.5)	2450	31.572	1.156	0.599
10 Vector Strength	$\bar{a}$	lines 4, 9	0.384	0.232	0.096	0.050
11 Parameter	k	Table II	0.832	0.477	0.194	0.100
12 Factor	$\sqrt{n/12}$	line 4	23.509	3.367	1	1
13 Table			in graph	7	8	—
15 Graph			12	13	14	15

If the deviations between the observations and a circular uniform distribution are small, a criterion for the reality of a cycle is needed, see [6]. We want to know whether a mode and anti-mode exist or are spurious and due to chance. In statistical terms this means: let the population value of the parameter be κ (e.g. zero); what is the probability of obtaining a certain value of k or a larger one from a sample of

GRAPH 15
PIG IRON PRODUCTION, 1936 - UNIFORM SCALE

size n? A similar, but more complicated, problem is to test the differences between the modes and values of the parameter k obtained from two samples. Finally a test for the goodness of fit between theory and observations is needed.

The following systematic deviations between theory and observations may exist: the observed values about the mode exceed (fall short

of) the theoretical values and show the opposite behavior at the antimode; the observed mode differs sensibly from the theoretical one; the observed antimode is not distant 180° from the observed mode. Finally there may be two modes and two antimodes. One of the latter may not be visible. In all these cases the chance distribution about one fixed date does not explain the observations. Then there must exist systematic reasons which have to be investigated from the intrinsic nature of the observations.

The simplest statistical explanation for such divergencies is the assumption that two cycles are involved instead of one. Two uniform distributions give again a uniform distribution. The composition of a uniform and a circular normal distribution leads, as shown before, to a circular normal distribution. The composition of two circular normal distributions with the same values α and k leads again to a circular distribution with α and k. Finally let $\alpha_1 \neq \alpha_2$ be the two modes, let $k_1 \neq k_2$ be the two measures of concentration, let A_1 and A_2 be the relative weights of the two constituents, then the composite distribution is

$$\phi(\alpha) = \frac{A_1 e^{k_1 \cos(\alpha - \alpha_1)}}{2\pi I_0(k_1)} + \frac{A_2 e^{k_2 \cos(\alpha - \alpha_2)}}{2\pi I_0(k_2)}.$$

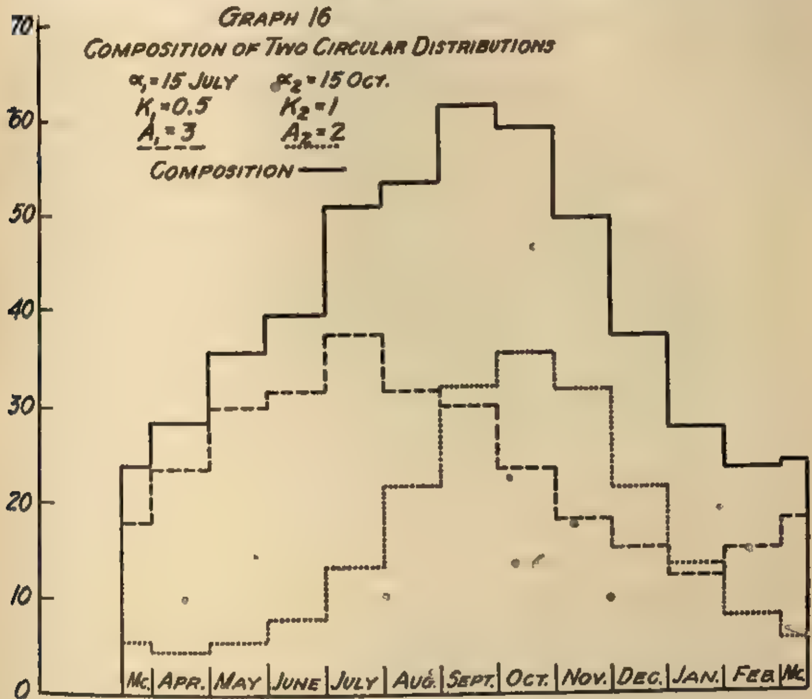
If the two modes of the component distributions are sufficiently near to each other, the combination may lead to an apparent single mode situated somewhere between them. If the two modes are sufficiently apart, the composition may lead to one mode and one hump, corresponding to an inflection point in linear scale, or to two modes, shifted of course from the original modes. The antimode or antimodes are no longer at a distance of 180 degrees from the mode or modes. The location of the modes or antimodes formally obtained by differentiation of $\phi(\alpha)$ are complicated functions of α_1, α_2, k_1, k_2 and A_1, A_2.

Graph 16 obtained from Table 1 and traced to linear scale, shows two component distributions with modes α_1 and α_2 at 15 July and 15 October, with parameters $k_1 = .5$ and $k_2 = 1$ and weights $A_1 = 3$, $A_2 = 2$. They combine into an asymmetrical distribution with a mode in September and a minimum in February.

Graph 17 traced in aequiareal scale shows the composition of two circular distributions of equal weight and with values $k = 2$, centered on $\alpha_1 = 0$ $\alpha_2 = \pi$ together with the two components.

The converse problem of separating an observed asymmetrical or symmetrical circular distribution into two symmetrical parts presents

no logical difficulties. However, great analytic difficulties arise from the estimation of the five parameters α_1, α_2, k_1, k_2, where $A_1=A$, $A_2=1-A$. Karl Pearson [13] has solved the corresponding problem within the linear normal distribution. In the general case it leads to an equation of the ninth degree. The difficulties are reduced see [3], if certain assumptions can be made about the modes or the values of the parameters k or the weights. By this procedure asymmetrical



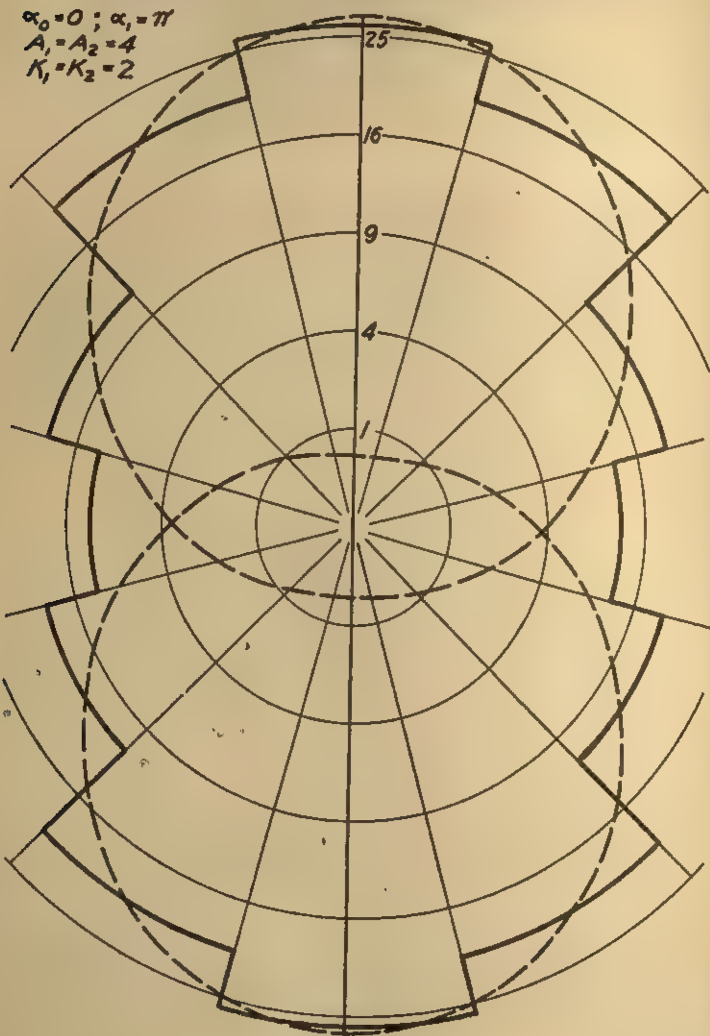
periodic phenomena can be analyzed by the addition of circular normal distributions.

Another method consists of wrapping an asymmetrical unlimited distribution around the circle. Even after the above problems are solved there will remain many cycles that cannot be analyzed by circular distributions. Statistics were not statistics if everything could be explained by this method.

Acknowledgment: The author is indebted to Arnold Court (University of California, Berkeley), David Durand (National Bureau of Economic Research), and J. A. Greenwood (Manhattan Life Insurance Co.) for valuable suggestions.

GRAPH 17

COMPOSITION OF TWO CIRCULAR DISTRIBUTIONS



BIBLIOGRAPHY

[1] *Automobile Facts and Figures;* Automobile Manufacturers Association, 31st edition, New York (1951), 5.

[2] Conrad, V. A., *Methods in Climatology*, Cambridge, Mass. (1946), chapter VIII.

[3] Court, A., "Separating frequency distributions into two normal components," *Science*, 110 (1949), 24.

[4] Court, A., "Some new statistical techniques in geophysics," *Advances in Geophysics*, edited by H. L. Landsberg, vol. 1, New York (1952), 75–83.

[5] Croxton, F. F., and Cowden, D. J., *Applied General Statistics*, New York (1940), 374.

[6] Durand, David, and Greenwood, J. A., "The integral solution of Pearson's random walk problem and related matters," *Annals of Mathematical Statistics*, 24 (1953), 686–87.

[7] Dyck, H. D., and Mattice, W. A., "A study of excessive rainfalls," *Monthly Weather Review*, 69 (1941), 293–302.

[8] Gumbel, E. J., Greenwood, J. A., and Durand, David A., "The circular normal distribution: theory and tables," *Journal of the American Statistical Association*, 48 (1953), 131–52.

[9] *Hydrology Handbook*, American Society of Civil Engineers, New York, (1949), 127.

[10] Kendall, M. G., *The Advanced Theory of Statistics*, London (1946), 127.

[11] Kuznetz, Simon, *Seasonal Variations in Industry and Trade*, National Bureau of Economic Research, New York (1933), 368.

[12] Pearl, R., *Introduction to Medical Biometry and Statistics*, Philadelphia (1940), 153.

[13] Pearson, K., "On the dissection of frequency curves," *Philosophical Transactions*, Royal Society, 185 A (1894), 71–110.

[14] Thompson, R. W. S., "The application of statistical methods in the determination of the yield of a catchment from run-off data," with a statistical note by D. H. Thomson, *Journal of the Institution of Water Engineers*, 4 (1950), 397.

A NEW TYPE OF CONTROL CHART LIMITS FOR MEANS, RANGES, AND SEQUENTIAL RUNS

H. Weiler
New South Wales University of Technology, Sydney, Australia

Since most conventional control chart limits are designed so that the ratio of the expected number of false alarms to the number of samples tested is fixed in advance irrespective of the sample size, the ratio of false alarms to the number of *articles* tested varies with the sample size. In this paper, control charts are designed for which the expected ratio of false alarms to the number of articles tested is independent of the sample size. The power of the control charts in relation to sample size is investigated.

1. INTRODUCTION AND SUMMARY

The usual control chart controlling the mean of a normal population is constructed in the following way: After the mean and standard deviation of the population have been reliably estimated, samples of fixed size n are selected and their arithmetic means $\bar{x} = \Sigma\, x/n$ are calculated. A chart is then constructed with control limits $m \pm B\sigma/\sqrt{n}$, where m and σ are estimates of the population mean and standard deviation, and B a constant. The various values of $\bar{x}$ are entered in the chart in chronological order, and as soon as one such value falls outside the control limits, production is stopped to allow investigation.

It is customary to take B equal to 3 or 3.09, irrespective of the sample size n. If $B = 3.09$ and if the population mean and standard deviation remain unchanged, an average of 500 samples will be required to produce one $\bar{x}$ value above the upper or below the lower limit. This means that an average of $500n$ articles will be tested for every one false alarm raised. The usual practice of setting control limits entails therefore that the average number of articles tested between two false alarms depends on the sample size n used for the control chart. This is likely to be an undesirable feature in quality control, where the cost of inspection is usually proportional to the number of articles inspected. The production engineer will therefore be interested in the number of articles rather than the number of samples tested.

In this paper, tables are provided for the determination of control limits for which the average number of false alarms (or Type I errors) is a fixed percentage of the number of articles tested, *independent of the sample size*. Curves similar to power curves are drawn for various sample sizes, giving the average amount of inspection in terms of the

amount by which the population mean has changed. A similar procedure is adopted for range charts.

One important result of the investigations is that the power of a mean chart increases rapidly with the sample size. For instance, the average amount of inspection required to detect a given change of the population mean is in most practical cases about twice as large for sample size $n=5$ as it would be if a chart for sample size $n=10$ were used (unless the change of the mean is very large). For samples of 20, the average amount of inspection can, under certain favorable circumstances be as little as one quarter of the amount required with samples of 5. This has already been pointed out in two previous papers [2, 3]; the results obtained here are even more striking.

However, since it is often necessary to maintain small samples in spite of the loss of power, it was shown in [3] that the power of small samples can be improved by charts using sequential runs. In the present paper, tables similar to those mentioned above are provided for the determination of suitable control limits for run charts. The corresponding curves show that the amount of inspection is greatly reduced, although it is frequently still high compared with the amount required for simple large sample charts.

Finally, run charts for ranges are briefly investigated. It turns out that the power of range charts is practically independent of the sample size and that the use of runs does not represent an improvement.

2. CONTROL LIMITS FOR MEAN CHARTS

If P is the probability that a random sample mean falls above the upper control limit, then about $100P$ samples in every 100, or one in every $1/P$ samples will fall above the upper control limit. It follows that on the average n/P *articles* will be tested before an alarm is raised at the upper limit.

Now, let p be the probability that a random sample causes a *false* alarm at the upper limit, and let a be the average number of articles tested before the false alarm is raised. We have then $a=n/p$ or $p=n/a$. If n and a are given, p can be determined; and if we assume x to be normally distributed, we can determine B such that the probability is p that $\bar{x}$ exceeds the upper control limit $m+B\sigma/\sqrt{n}$, where m is the mean and σ the standard deviation of the parent population. If, for instance, we take $n=5$ and $a=5000$, we have $p=0.001$, and we obtain $B=3.09$ from a set of normal tables.

Proceeding in this manner for various sample sizes, but keeping $a=5000$ fixed, we obtain the values of p, B, $B/\sqrt{n}$ shown in Table I.

TABLE I

VALUES OF p, B, $B/\sqrt{n}$, WHEN $a=5000$

Notation: p = probability that a random sample causes a false alarm; n = sample size; a = average number of articles tested before an alarm occurs at the upper (or lower) control limit, if the process remains under control.

n	3	4	5	6	8	10
p	.0006	.0008	.0010	.0012	.0016	.0020
B	3.24	3.16	3.09	3.03	2.95	2.88
$B/\sqrt{n}$	1.87	1.58	1.38	1.24	1.04	0.91

Similar arguments hold for the lower control limit, so that in each case an average of one false alarm above the upper and one false alarm below the lower control limit must be expected for every 5000 articles tested. In other words, we will be able to test an average of 2500 articles before one false alarm is raised either way.

Control limits for values of a other than 5000 may be calculated in the same way. Table II gives values of $B/\sqrt{n}$ for various values of a and for sample sizes ranging from 3 to 50.

TABLE II

VALUES OF $B/\sqrt{n}$ FOR CONTROL LIMITS $m \pm (B/\sqrt{n})\sigma$ FOR MEAN CHARTS

Notation: n = sample size; m = population mean; σ = population standard deviation; a = average number of articles tested before an alarm occurs at the upper (or lower) control limit, if the process remains under control.

n	$a=1000$	$a=2000$	$a=3000$	$a=4000$	$a=5000$
3	1.59	1.71	1.78	1.83	1.87
4	1.33	1.44	1.50	1.54	1.58
5	1.15	1.26	1.31	1.35	1.38
6	1.02	1.12	1.17	1.21	1.24
7	0.93	1.02	1.07	1.10	1.13
8	0.85	0.94	0.99	1.02	1.04
9	0.79	0.87	0.92	0.95	0.97
10	0.74	0.82	0.86	0.89	0.91
15	0.560	0.628	0.665	0.690	0.710
20	0.459	0.520	0.553	0.577	0.593
25	0.392	0.448	0.479	0.500	0.515
30	0.343	0.396	0.425	0.444	0.458
35	0.306	0.356	0.383	0.402	0.416
40	0.277	0.324	0.350	0.368	0.381
45	0.253	0.299	0.324	0.340	0.353
50	0.232	0.277	0.301	0.317	0.329

To determine the control limits $m \pm (B/\sqrt{n})\sigma$, we estimate the mean m and standard deviation σ of the population and obtain the appropriate factor $B/\sqrt{n}$ from Table II. If, for instance, we want to use samples of size $n=8$, and if we wish to test an average of 2000 articles before raising a false alarm either way ($a=4000$), we find $B/\sqrt{n}=1.02$, and the control limits are $m \pm 1.02\sigma$.

3. THE AVERAGE AMOUNT OF INSPECTION WHEN THE POPULATION MEAN CHANGES

If the population mean changes from m to $m+k\sigma$ ($k>0$) while σ remains constant, the variate $\bar{x}$ will have the mean $m+k\sigma$ and the standard deviation $\sigma/\sqrt{n}$, so that the variate

$$z = \frac{\bar{x} - (m + k\sigma)}{\sigma/\sqrt{n}} \tag{1}$$

is a standardized normal variate (mean zero and S.D. one). The probability that $\bar{x}$ exceeds the upper control limit $m+B\sigma/\sqrt{n}$ is then

$$\begin{aligned} P &= \Pr\left\{\bar{x} \geqq m + \frac{B\sigma}{\sqrt{n}}\right\} = \Pr\left\{\bar{x} - m - k\sigma \geqq \frac{B\sigma}{\sqrt{n}} - k\sigma\right\} \\ &= \Pr\left\{\frac{\bar{x} - m - k\sigma}{\sigma/\sqrt{n}} \geqq B - k\sqrt{n}\right\} = \Pr\left\{z \geqq B - k\sqrt{n}\right\}. \end{aligned} \tag{2}$$

The average number of articles tested before an alarm is raised at the upper limit is then $A(n)=n/P$, where

$$P = \frac{1}{\sqrt{2\pi}} \int_{B-k\sqrt{n}}^{\infty} \exp\left(-\tfrac{1}{2}z^2\right) dz. \tag{3}$$

For any given n and B, the average number of articles tested (or "average amount of inspection") $A(n)$ is a function of k, whose values can be calculated with the aid of a set of normal tables.

In Charts I and II, $A(n)$ is plotted against k for various values of n, while $a=2000$ and $a=5000$, respectively. Both charts show clearly the superiority of large sample sizes over a wide range of k values. For instance, $n=10$ is more powerful than $n=5$ for any value of k less than 1.2 or 1.3, and $n=20$ is more powerful than $n=5$ for k less than 1.0 or 1.1. In particular, for $a=5000$ and $k=0.7$, the average amount of inspection is 80 for $n=5$, 40 for $n=10$, and only 30 for $n=20$. The saving of inspection is even greater when samples of 50 are used, but the range of k for which such a large sample size is powerful is rather restricted ($k<0.5$, say). Only for the detection of very large shifts of the mean will small samples be more powerful than large samples.

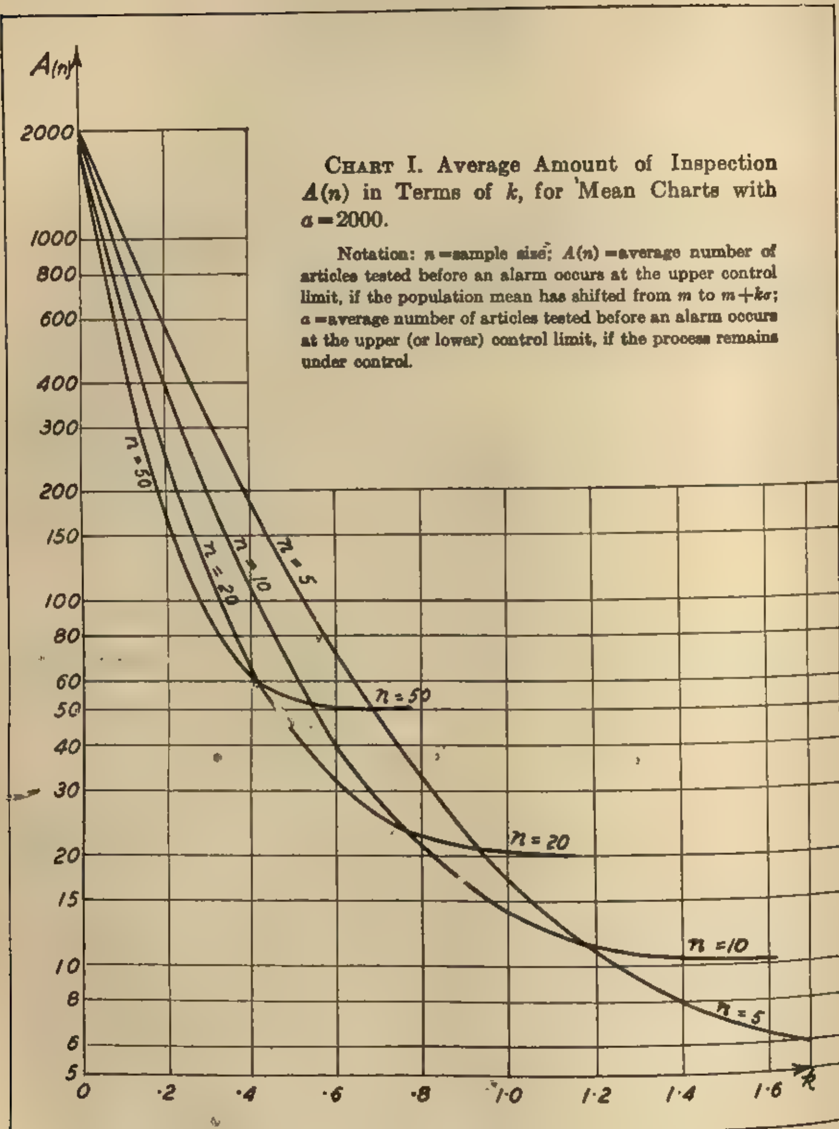


CHART I. Average Amount of Inspection $A(n)$ in Terms of k, for Mean Charts with $a=2000$.

Notation: n = sample size; $A(n)$ = average number of articles tested before an alarm occurs at the upper control limit, if the population mean has shifted from m to $m+k\sigma$; a = average number of articles tested before an alarm occurs at the upper (or lower) control limit, if the process remains under control.

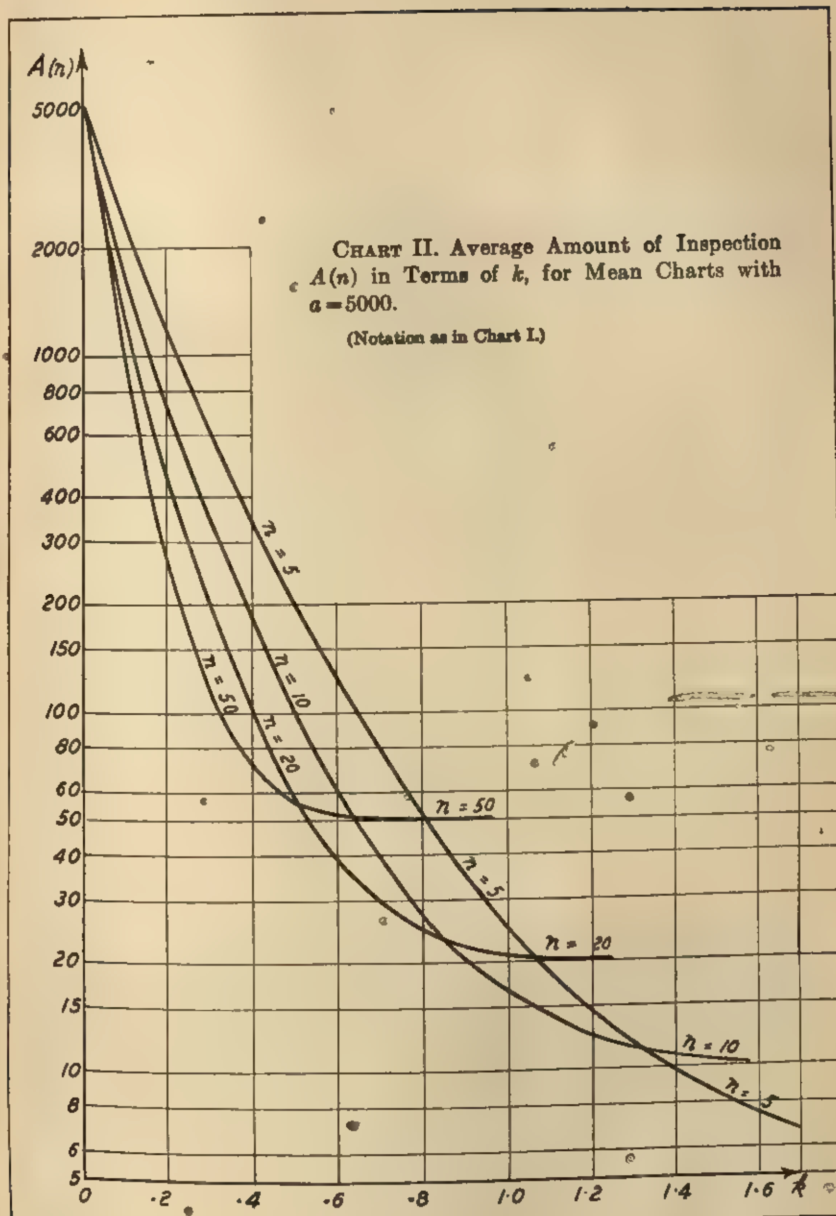


CHART II. Average Amount of Inspection $A(n)$ in Terms of k, for Mean Charts with $a=5000$.

(Notation as in Chart I.)

In Chart III, $A(n)$ is plotted against k for the fixed sample size $n=10$ and various values of a. They show that we are able to reduce the average amount of inspection by making a smaller, that is, by increasing the average rate of false alarms. The curves show, however, that small values of a are useful only for the detection of small changes of the population mean. If, for instance, $a=5000$ is replaced by $a=1000$, the average number of false alarms becomes 5 times as large, but the number of real alarms when $k=0.6$ is only doubled.

With the help of Charts I and II, it will be easy to decide whether large or small samples should be taken. If we are mainly concerned with the detection of shifts of the mean larger than (say) one standard deviation, small samples are advisable. If, on the other hand, small values of k are expected, the samples should be large. If, for instance, k is expected to lie between 0.4 and 0.8, both charts show that the sample size $n=20$ is more powerful than $n=10$. If, on the other hand, k is expected to lie between 0.8 and 1.2 (say), samples of 10 are better than samples of 20. Once a and n are fixed, the control limits can be determined by means of Table II.

4. CONTROL LIMITS FOR RANGE CHARTS

If $x_1, x_2, \cdots, x_n$ is a random sample of n observations arranged in order of magnitude, the variate $R=x_n-x_1$ is called the range of the sample. Instead of the range R, we shall consider the variate $w=R/\sigma$, where σ is the standard deviation of the parent population.

No simple expression exists for the probability law $\phi_n(w)$ of w, but tables have been prepared for the probability integral

$$(4) \qquad p_n(W) = \int_0^W \phi_n(w)dw$$

when the parent population is normal [1]. This expression represents the probability that a random sample of size n has a range less than a given multiple W of the population standard deviation σ. The probability that the range exceeds the value $W\sigma$ is equal to $1-p_n(W)$.

Like the control chart for means, most control charts for ranges have an upper and a lower control limit. When a sample range falls outside the control limits, it is regarded as an indication that the standard deviation of the population has changed.

We shall determine the control limits $W_1\sigma \leq R \leq W_2\sigma$ such that the expected number of articles tested before a false alarm at the upper limit $W_2\sigma$, or at the lower limit $W_1\sigma$, is equal to a preassigned number a for any sample size n.

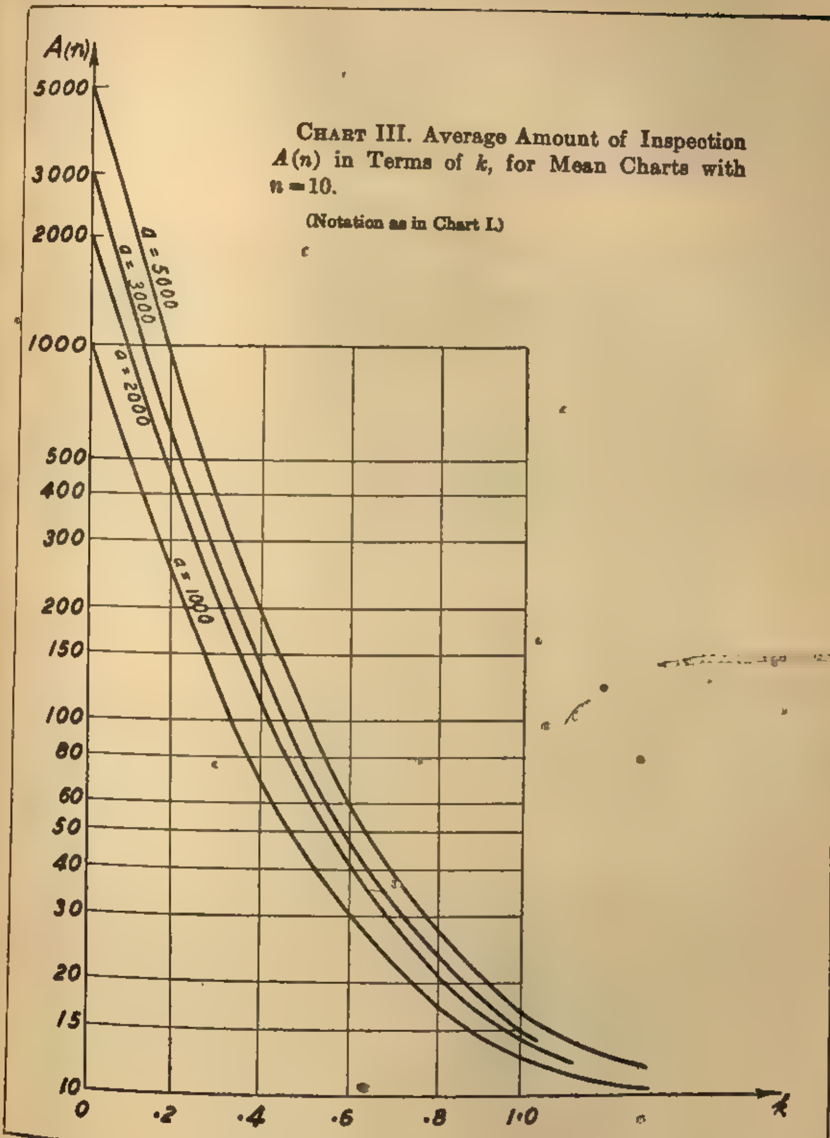


CHART III. Average Amount of Inspection $A(n)$ in Terms of k, for Mean Charts with $n=10$.

(Notation as in Chart I.)

Let $p_1 = p_n(W_1)$ and $p_2 = p_n(W_2)$ be the probabilities that the range of a random sample falls below the lower and above the upper limit, respectively. As in Section 2, we have then $p_1 = p_2 = n/a$. The values of W_1 and W_2 can then be determined easily by means of the above mentioned tables [1]. If, for instance, $n=5$ and $a=5000$, we have $p_1 = p_2 = 0.001$, and the tables provide $W_1 = 0.37$ and $W_2 = 5.48$. Similarly, the values of W_1 and W_2 may be found for other values of n and a; they are tabulated in Table III.

TABLE III

VALUES OF W_1 AND W_2 FOR CONTROL LIMITS $W_1\sigma \leqq R \leqq W_2\sigma$

Notation: n = sample size; R = sample range; σ = population standard deviation; $W_1\sigma$ = lower control limit; $W_2\sigma$ = upper control limit; a = average number of articles tested before an alarm occurs at the upper (or lower) control limit, if the process remains under control.

n	$a=1000$	$a=2000$	$a=3000$	$a=4000$	$a=5000$
3	4.64 0.10	4.00 0.06	5.06 0.06	5.18 0.05	5.25 0.04
4	4.78 0.31	5.05 0.25	5.23 0.22	5.31 0.20	5.40 0.18
5	4.88 0.55	5.15 0.46	5.30 0.42	5.41 0.39	5.48 0.37
6	4.96 0.78	5.23 0.67	5.38 0.62	5.48 0.58	5.55 0.56
7	5.02 0.93	5.29 0.86	5.44 0.80	5.54 0.76	5.63 0.73
8	5.07 1.16	5.34 1.04	5.47 0.97	5.58 0.93	5.67 0.90
9	5.12 1.33	5.38 1.19	5.53 1.13	5.64 1.09	5.70 1.05
10	5.16 1.47	5.42 1.33	5.57 1.27	5.67 1.22	5.73 1.18

5. THE AVERAGE AMOUNT OF INSPECTION WHEN THE POPULATION STANDARD DEVIATION CHANGES

Suppose that the standard deviation of the parent population changes from σ to $\sigma' = k\sigma$, $k>1$. The variate $w' = R/\sigma'$ has then the same distribution as previously the variate $w = R/\sigma$. It follows that

the probability that the range R of a random sample falls above the upper limit $W_2\sigma$ is

$$
\begin{aligned}
(5) \qquad P &= 1 - \Pr\{R \leq W_2\sigma\} = 1 - \Pr\left\{w' < W_2 \frac{\sigma}{\sigma'}\right\} \\
&= 1 - \Pr\left\{w' < \frac{W_2}{k}\right\} = 1 - \int_0^W \phi_n(w)dw,
\end{aligned}
$$

where $W = W_2/k$.

The average number of articles tested before a change of the standard deviation from σ to $k\sigma$ is detected is, as in Section 3, $A(n) = n/P$. $A(n)$ depends on n and k and also on the number a defined in Section 4. Its relation to k for any given values of a and n can be easily obtained by means of the range tables [1].

In Charts IV and V, $A(n)$ is plotted against k for some values of n and a. Chart IV shows that little is gained by increasing the sample size from $n=5$ to $n=10$. Chart 5 shows that the amount of inspection may be reduced by making a smaller, but, as for mean charts, small values of a are useful only for the detection of small changes in the standard deviation.

6. CONTROL LIMITS FOR MEAN CHARTS USING RUNS

It was shown in [3] that the power of mean charts for small samples can be improved by the use of runs. For such charts, control limits $m \pm B\sigma/\sqrt{n}$ are determined, and as soon as λ successive $\bar{x}$ values fall above the upper or below the lower control limit, alarm is raised and production is stopped to allow investigation. We shall again determine B such that the average number of articles tested between two successive false alarms is independent of the sample size.

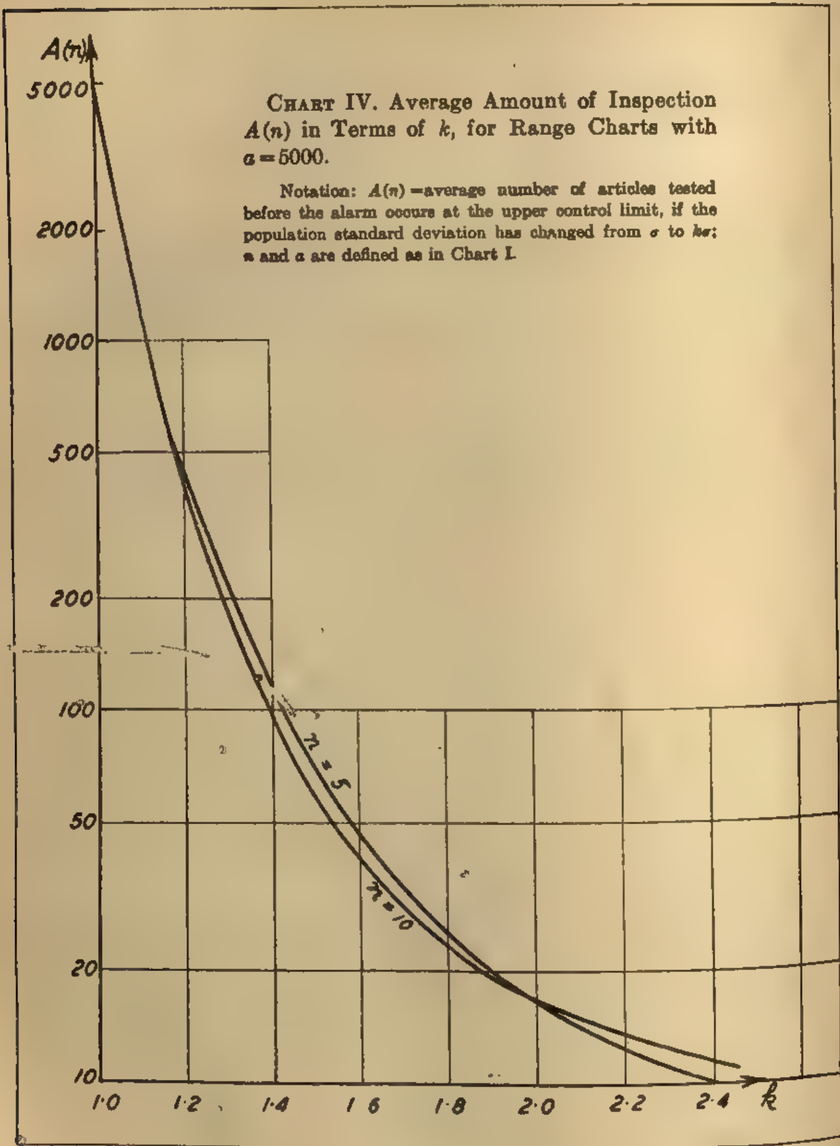
If P is the probability that a random $\bar{x}$ value falls above the upper control limit, the average number of samples that will pass before λ successive $\bar{x}$ values fall above the upper limit is [3] $S = P^{-1} + P^{-2} + \cdots + P^{-\lambda}$. The values of P for any B, k, and n are again given by equation (3). The average number of articles tested before an alarm is raised at the upper control limit, is then $A(n) = nS$.

In particular, a *false* alarm is raised when $k=0$. The probability P then becomes

$$
(6) \qquad p_0 = \frac{1}{\sqrt{2\pi}} \int_B^\infty \exp\left(-\tfrac{1}{2}z^2\right)dz,
$$

CHART IV. Average Amount of Inspection $A(n)$ in Terms of k, for Range Charts with $a=5000$.

Notation: $A(n)$ = average number of articles tested before the alarm occurs at the upper control limit, if the population standard deviation has changed from σ to $k\sigma$; n and a are defined as in Chart I.



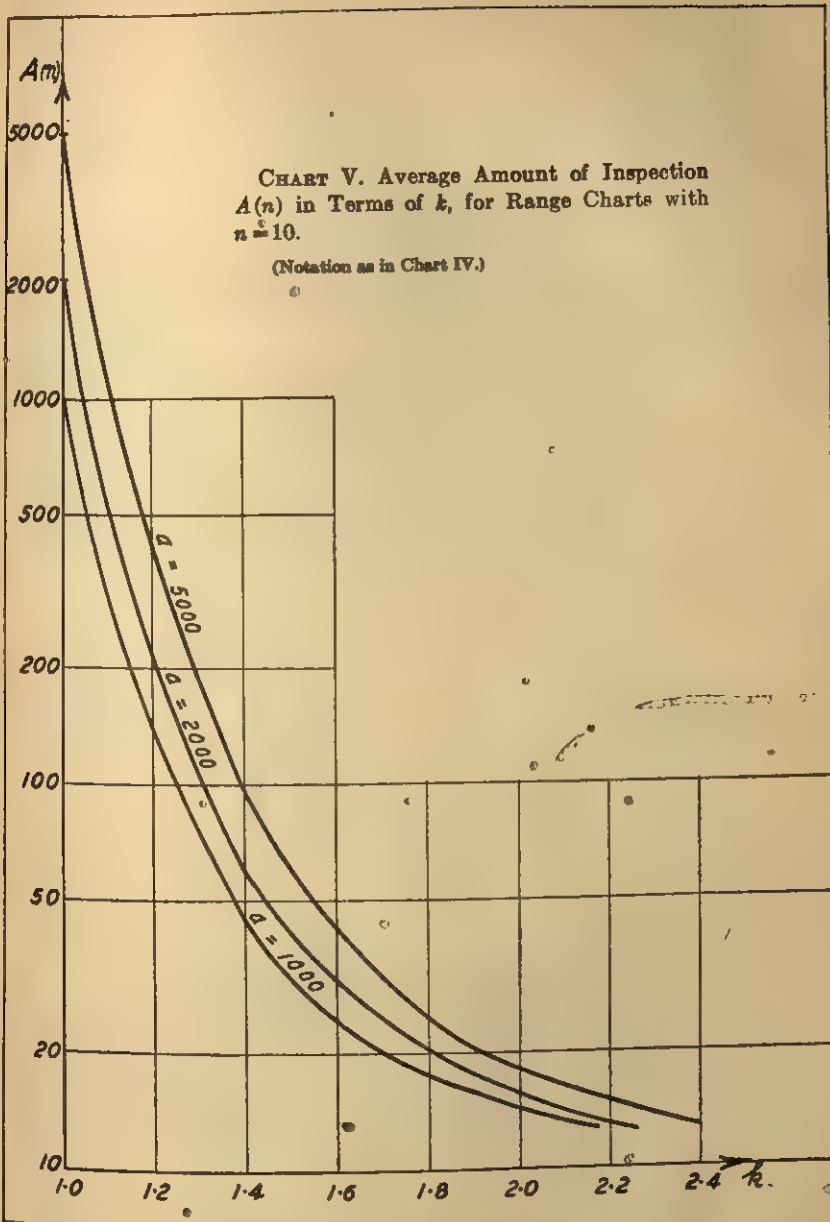


CHART V. Average Amount of Inspection $A(n)$ in Terms of k, for Range Charts with $n = 10$.

(Notation as in Chart IV.)

and the average number of articles tested before a false alarm is raised at the upper control limit is $a=ns$, where $s=p^{-1}+p^{-2}+\cdots+p^{-\lambda}$. If, for instance, $a=5000$, $\lambda=2$, and $n=4$, we find $p=0.0287$. A set of normal tables supplies $B=1.90$, and $B/\sqrt{n}=0.95$ follows. In this way we can determine control limit factors $B/\sqrt{n}$ for various values of λ, a, and n, as shown in Table IV.

TABLE IV

VALUES OF $B/\sqrt{n}$ FOR CONTROL LIMITS $m\pm(B/\sqrt{n})\sigma$ FOR RUN CHARTS

Notation: n = sample size; m = population mean; σ = population standard deviation; a = average number of articles tested before a run of λ sample means occurs above the upper (or below the lower) control limit, if the process remains under control.

λ	n	$a=1000$	$a=2000$	$a=3000$	$a=4000$	$a=5000$
2	1	1.85	2.00	2.09	2.15	2.20
	4	0.76	0.85	0.90	0.93	0.95
	5	0.65	0.73	0.78	0.81	0.83
	8	0.47	0.53	0.57	0.60	0.62
	10	0.40	0.46	0.49	0.52	0.53
	20	0.23	0.28	0.31	0.33	0.34
3	1	1.26	1.39	1.47	1.52	1.56
	4	0.49	0.56	0.59	0.63	0.65
	5	0.41	0.48	0.52	0.55	0.56
	8	0.33	0.34	0.37	0.40	0.41
	10	0.23	0.29	0.32	0.34	0.35
4	1	0.89	1.01	1.08	1.13	1.16
	4	0.31	0.38	0.42	0.45	0.47
	5	0.25	0.32	0.35	0.38	0.40
	8	0.16	0.22	0.25	0.27	0.28
	10	0.12	0.18	0.21	0.22	0.24

7. THE AMOUNT OF INSPECTION FOR RUN CHARTS CONTROLLING THE MEAN

In this section, we shall discuss power curves similar to those given in [3]. The curves shown in this paper differ from those in [3] only in the values of B, which here are functions of n such that the average number a of articles tested between two successive false alarms becomes independent of the sample size.

To find the value of $A(n)=nS$ for any given n, λ, a, and k, we determine first the value of $B/\sqrt{n}$ from Table IV. We then find P by equation (3), using a set of normal tables, and deduce $S=P^{-1}+P^{-2}$

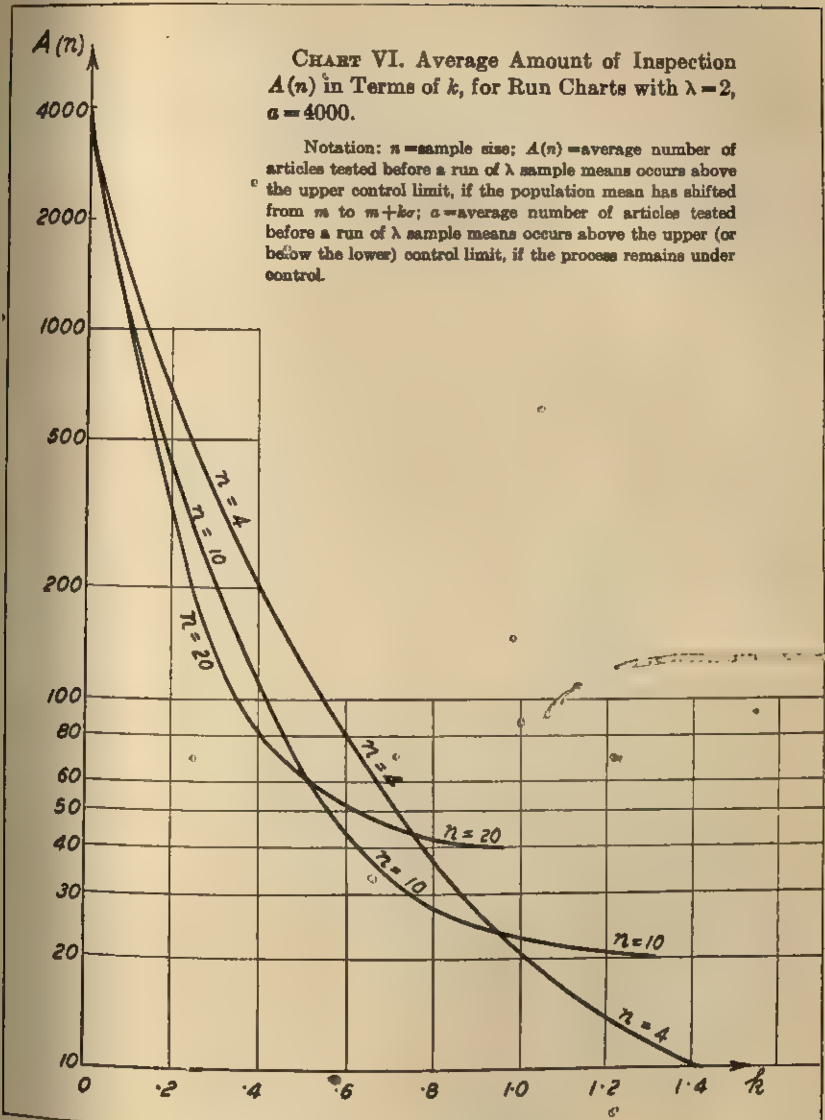


CHART VI. Average Amount of Inspection $A(n)$ in Terms of k, for Run Charts with $\lambda = 2$, $a = 4000$.

Notation: n = sample size; $A(n)$ = average number of articles tested before a run of λ sample means occurs above the upper control limit, if the population mean has shifted from m to $m + k\sigma$; a = average number of articles tested before a run of λ sample means occurs above the upper (or below the lower) control limit, if the process remains under control.

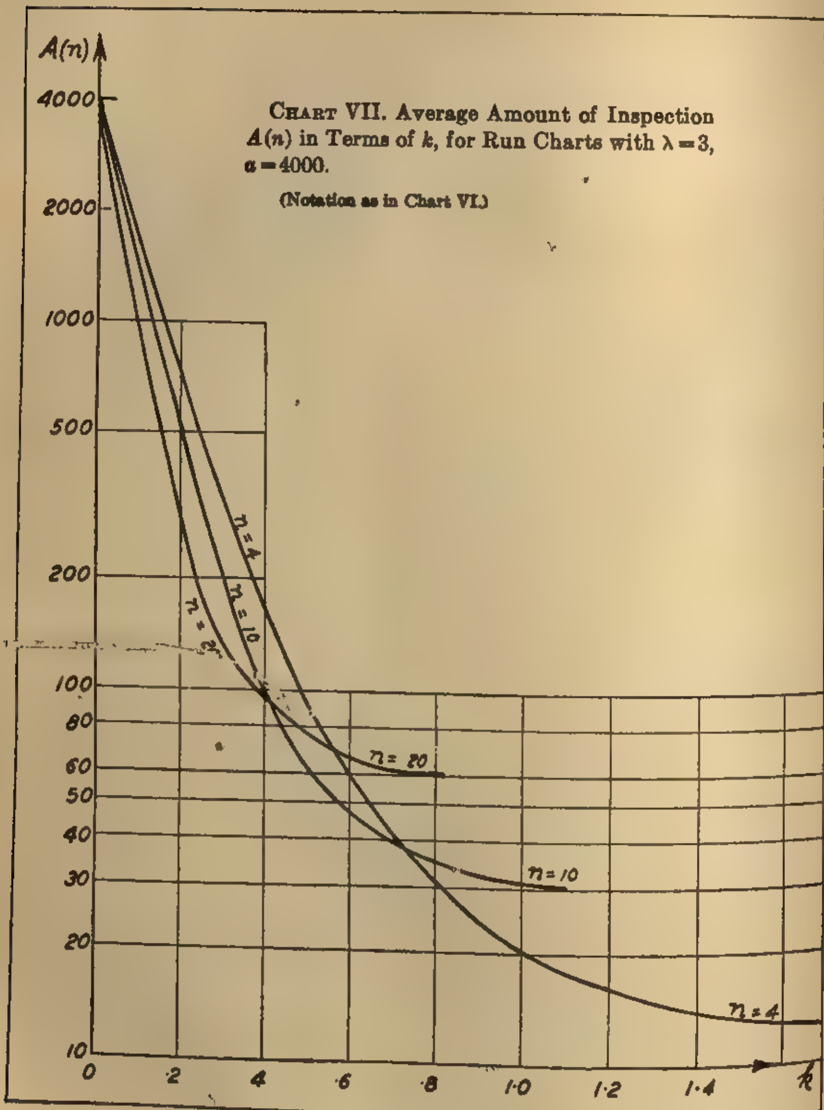


CHART VII. Average Amount of Inspection $A(n)$ in Terms of k, for Run Charts with $\lambda=3$, $a=4000$.

(Notation as in Chart VI.)

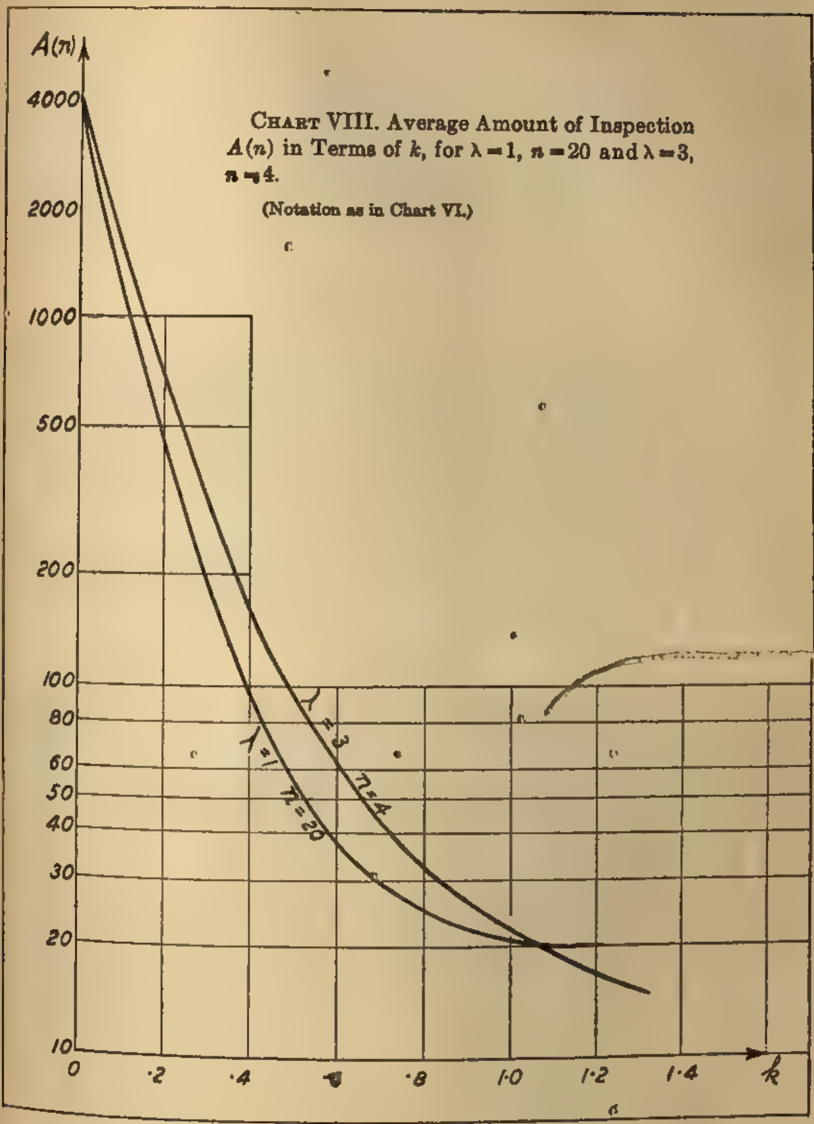


Chart VIII. Average Amount of Inspection $A(n)$ in Terms of k, for $\lambda = 1$, $n = 20$ and $\lambda = 3$, $n = 4$.

(Notation as in Chart VI.)

$+\cdots+P^{-\lambda}$. (Actually, to plot the curves, we found it more convenient to start with given values of P and to deduce corresponding values of k, S, and $A(n)$.)

Chart VI shows that for $\lambda=2$ it is still advantageous to take a large sample size, such as $n=10$. On the other hand, $n=20$ represents only a slight improvement on $n=10$, and this only over a restricted range for $k(k<0.6)$. When $\lambda=3$ is used (Chart VII), the advantage of larger samples becomes even less pronounced. However, Chart VIII shows that there is still a considerable saving in inspection when a simple chart ($\lambda=1$) for large samples can be used instead of a run chart for small samples.

8. THE USE OF RUNS FOR RANGE CHARTS

Instead of stopping the production when a single value of R falls outside the control limits R_1, R_2 of an ordinary range chart, we may calculate a pair of narrower control limits R_1', R_2', and stop production as soon as λ successive R values fall above the upper or below the lower control limits. The new limits are obtained as in Sections 4 and 6 (e.g. for $n=5$, $\lambda=2$, we obtain $R_2'=4.08\sigma$).

Power curves can then be plotted by the methods described in Sections 5 and 7, but results obtained show that the use of runs reduces rather than improves the power of the chart.

9. REFERENCES

[1] Pearson, E. S., and Hartley, H. O., "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, 32 (1942), 301–8.

[2] Weiler, H., "On the most economical sample size for controlling the mean of a population," *Annals of Mathematical Statistics*, 23 (1952), 247–54.

[3] Weiler, H., "The use of runs to control the mean in quality control," *Journal of the American Statistical Association*, 48 (1953), 816–25.

PROCEEDINGS

AMERICAN STATISTICAL ASSOCIATION
113TH ANNUAL MEETING

SHOREHAM HOTEL, WASHINGTON, D. C.
DECEMBER 28, 1953

MINUTES OF THE ANNUAL BUSINESS MEETING

The meeting was called to order by William G. Cochran, outgoing President of the Association.

Report of the Committee on Elections

A Report of the Committee on Elections shows the following officers elected for 1954:

President Elect	Ralph J. Watkins
Vice President (1954–56)	Henry Scheffé
Directors (1954–56)	Jacob Marschak
	Donald C. Riley
Representative at Large (1954–55)	Daniel B. deLury
District Representatives	
Northeastern District	Chester I. Bliss
Eastern District	George Garvy
Southeastern District	Frank A. Hanna
North Central District	Paul R. Rider, Lucile Derrick
South Central District	William Kester
Western District	John C. McKee

Report of the Board of Directors for 1953

The Report of the Board of Directors was read and accepted. The Report is published separately following the Minutes of this Meeting.

Report of the Secretary-Treasurer for 1953

Samuel Weiss read the Report of the Secretary-Treasurer for the year 1953. The Report was accepted. The Secretary-Treasurer's Report is published separately following the Minutes of this Meeting.

Report of the Committee on Resolutions

The Committee on Resolutions presented the following to the membership for their consideration:

1. Resolution regarding the Program Committee
   RESOLVED that the members and officers of the American Statistical Association express deep appreciation for the excellent program prepared by members of the Program Committee under the leadership of Herbert Solomon, Chairman.
2. Resolution regarding the Local Arrangements Committee and the Washington, D. C. Chapter.

RESOLVED that the members and officers of the American Statistical Association express their profound appreciation to the Local Arrangements Committee under the Chairmanship of Donald C. Riley and to all of the individuals of the Washington, D. C. Chapter for their outstanding work and hospitality in connection with the arrangements for the 113th Annual Meeting of the Association.

The resolutions were approved.

Report on the Schedule of Forthcoming Meetings

The forthcoming meetings of the Association are scheduled as follows:

1954—Annual Meeting—Montreal, Canada—September 10–13, 1954
1954—Regional Meeting—San Francisco Regional Conference—Berkeley, California—December 1954
1955—Annual Meeting—New York City—December 27–29, 1955

There being no new business, the meeting was adjourned.

REPORT OF THE BOARD OF DIRECTORS, 1953

Activities of the Association during 1953 were vigorous and widespread in scope. This was coupled with a most successful financial year. Details of finance and membership will be found in the Secretary-Treasurer's Report.

1. *Sections and Committees*

The Board has granted sectional status to the Social Statistics Section, formerly the Committee on Statistics in the Social Sciences. This Committee requested sectional status at the Meeting of the Incoming Board and Council in December, 1952. The matter was referred to the Committee on Committees, which, after reviewing the final charter of the proposed section, recommended to the Board that approval be given. The approval of this Charter makes this the fourth Section of the Association; the other three being the Biometric Section, the Business and Economics Statistics Section and the Section on the Training in Statistics.

These four Sections and the Committee on Statistics in the Physical Sciences have been extremely active in the formulation and planning of the Annual Meeting programs, as well as the planning for a number of successful Regional Meetings. Without any question, the formation of the Sections has been extremely beneficial to the Association. The variety of interests of the membership of the Association is best served by the activity of these Sections.

Two *ad hoc* committees were constituted by the Board this year. They are the *Ad Hoc* Committee on Publications Policy and the *Ad Hoc* Committee on Statistical Standards and Organization. Both groups were asked to make their Reports to the Board so that action may be taken on their recommendations in these two important areas.

The Board wishes to commend the 1953 Program and Local Arrangement Committees. The former, with Herbert Solomon as Chairman, has done an outstanding job of presenting a well-balanced, varied group of sessions for the Annual Meeting. The Local Arrangement Committee, chaired by Donald Riley, has worked very hard to ensure a successful meeting. Their efforts for ASA, in cooperation with the other societies meeting jointly, have produced a memorable convention.

2. Abstracts

This year for the first time the Association has published abstracts of papers presented at the previous Annual Meeting in the *Journal of the American Statistical Association*. The editor of these abstracts was Arman Alchian. The Board is very pleased to announce that Professor Alchian has agreed to continue in this position for the coming year. It is hoped that, with cooperation from persons presenting papers at this Annual Meeting, the abstracts will appear in an early issue of the 1954 volume of the *Journal*.

3. New Constitution

The new Constitution, approved by the membership in a mail ballot in 1953, will go into effect January 1, 1954. The changes incorporated into the new version are, for the most part, procedural rather than substantive, and should facilitate the operation of the Association. The draft version, which was ratified by the membership, was printed in the June–July, 1952 issue of the *American Statistician*.

4. New Chapters

During 1953 four new Chapters were granted Charters by the Board of Directors. The new groups are located in Tulsa, New Orleans, Puerto Rico and Milwaukee. The Board extends a welcome to these new Chapters.

The total number of active Chapters has now reached 31. Chapter meetings, often in cooperation with local chapters of other societies, offer a wide variety of topics to the membership in the fields of statistical interest.

5. Conferences and Meetings

During 1953 the Business and Economic Statistics Section sponsored two regional conferences. The first was held on April 30 and May 1 in cooperation with the Graduate School of Industrial Administration at the Carnegie Institute of Technology and the Pittsburgh Chapter of the American Marketing Association and the American Society for Quality Control. The Conference was devoted to modern statistical methods in business and industry.

The second Conference, sponsored jointly with the Wharton School of the University of Pennsylvania, within its theme of business statistics was divided into three main sessions—on capital outlays, production scheduling and sales forecasting. This Conference took place in Philadelphia on June 11 and 12. Both Conferences were well attended and very successful.

6. New Monograph

Following the Association's policy of a more vigorous program of publication, the Board has voted to publish as a monograph the complete Report of the ASA Committee to Advise the National Research Council Committee for Research in the Problems of Sex. This Report, dealing with the volume on the human male entitled "Sexual Behavior in the Human Male," by Professor Kinsey and his associates, will be available in 1954. Certain sections of the Report are also being published as articles in the *Journal of the American Statistical Association*.

7. New Appointments of ASA Representatives

Samuel S. Wilks was elected by the Council to continue to serve as the ASA Representative to the Social Science Research Council for a three-year term.

The National Bureau of Economic Research announced the resignation, due to illness, of Frederick C. Mills as ASA Representative to its Board of Directors. The Board has appointed W. Allen Wallis to complete the unexpired portion of Dr. Mills' term. A resolution was received from the National Bureau of Economic Research which expressed deep regret for the necessity of Dr. Mills' resignation and thanking the ASA for his valuable services during his tenure of office.

8. *Journal of the American Statistical Association*

The Board approved an increase in the budget for the *Journal* for 1953, with the result that more articles were published in the 1953 volume, and the number of pages increased by approximately 25 per cent over 1952. In addition, the Board approved an increase in the funds for editorial assistance, which heretofore has been supplied by the University of Chicago.

9. *Reduction in Dues for Foreign Members*

The Board voted to reduce the dues for persons residing outside North America, beginning in 1954. The reduction from $8.00 to $5.00 yearly will make it easier for many foreign members in terms of dollar exchange. It is expected that the slight loss in income from this reduction will be more than offset by the increase in new members outside North America. A 1954 drive for new foreign members is planned.

10. *Future Meetings*

Future plans of the Association call for the 1954 Annual Meeting to be held in Montreal, Canada, on September 10–13. Two Regional Meetings are also being planned for 1954. The Chicago Chapter, in cooperation with other Chapters in its District, has begun work on a meeting to be held in the spring, probably during April. A Western Regional Meeting has been scheduled for San Francisco in December of 1954, in conjunction with the American Association for the Advancement of Science. Final dates and headquarters for this conference have not yet been chosen. Advance publicity will be issued for both of these Regional Meetings.

The Montreal Meeting was planned for September to leave Christmas week free for members who usually attend annual Meetings, and to attract those who do not attend late December Meetings. In this sense it is an experiment designed to guide the Association in the selection of other meeting dates. The widest understanding of the nature of the experiment will help its success.

REPORT OF THE SECRETARY-TREASURER, 1953

A drive for an increase in new members which started in 1952 continued actively in 1953. This, together with a policy of careful economy, has resulted in increasing the Association's surplus by a wider margin than ever before.

Income for 1953 was budgeted at $52,650. The actual total 1953 income is $60,377.31. This difference is due to a rise in income of most budgeted items, but primarily results from increases in receipts from membership dues, sales of publications and subscriptions. Expenses were budgeted at $50,612. The actual total 1953 figure is $50,433.83. All expenses have been kept close to budget level, and small savings have been made on a number of different items. Thus,

the Association shows an increase to surplus of $9,943.48 at the end of 1953. This brings the total surplus to over $26,000, about halfway to the goal of a surplus equal to one year's income.

For the second successive year the number of members has shown a significant increase. At the beginning of 1953 the membership totaled 4,655. The number of new members for 1953 was 639, and 25 others reinstated their membership. At the end of 1953 approximately 400 members have been dropped from the rolls because of resignation, death or non-payment of dues. Thus, the net membership growth for 1953 is 264, and the Association starts 1954 with a total of more than 4,900 members. This is the highest level of membership ever reached by the Association. It is expected that the drive for more foreign members, as mentioned in the Board of Directors' Report, will add substantially to the membership in 1954.

Subscriptions to the *Journal of the American Statistical Association* have also been increasing. At the end of 1952 there were 1,248 subscribers to the *Journal*, while at the end of 1953 the figure had risen to 1,356. This increase is expected to continue in 1954.

Financial Recommendations

The Report of the Treasurer, shown separately, emphasizes that 1953 was the fourth year in succession in which the Association has accrued surplus. The Board of Directors for the past two years has recommended that the surplus be increased until it equals the income of the Association for one year. At that time it is felt that the Association will be in a much stronger position to expand its activities on a much wider scale. With this in mind, the Treasurer has planned to budget between $2,000 and $3,000 per year for addition to surplus until this goal is reached. The proposed income for 1954 is budgeted at approximately $55,000, while expense has been calculated at $52,500, leaving $2,500, for addition to surplus. Income has been figured very conservatively, while expense has been approximated as closely as possible, with the expectation that the surplus may be somewhat larger than budgeted.

April 19, 1954

To the Board of Directors of
American Statistical Association.

I have examined the attached financial statements of American Statistical Association relating to the year ended December 31, 1953. My examination was made in accordance with generally accepted auditing standards and, accordingly, included such tests of the accounting records and such other auditing procedures as were considered necessary in the circumstances.

The recorded cash receipts for the year were traced in the deposits shown on the bank statements and the amounts for dues and subscriptions were tested with the membership and subscription records. The paid checks were inspected and related vouchers tested in support of cash disbursements for the year. The bank balances were reconciled with amounts reported directly to me by the depositaries and the cash on hand at December 31, 1953 was verified by inspection. I did not check the membership and subscription records in detail or make any independent verification of the inventory of old *Journals*, the office records of which are based, in part, on data assembled in prior years.

In accordance with a resolution of the Board of Directors, the expense incurred

in publishing a directory, distributed to the membership in 1951, is being spread over a three-year period although such costs would appear to be applicable primarily to the year 1951. The accounts for the year ended December 31, 1953 reflect a charge of $831.86, representing the allocated portion of the directory expense applicable to that period.

In my opinion, the accompanying statements present fairly the position of American Statistical Association at December 31, 1953, and the results of its operations for the year, in conformity with generally accepted accounting principles applied on a basis consistent, except as mentioned in the preceding paragraph, with that of the preceding year.

JAMES G. JESTER

AMERICAN STATISTICAL ASSOCIATION
BALANCE SHEET

Assets

	December 31, 1953	December 31, 1952
Cash in banks and on hand	$52,431.05	$39,503.14
Accounts receivable	2,783.14	1,099.58
Investment in United States Savings Bonds, Series G, due 1962, at cost	3,100.00	3,100.00
Inventory of old *Journals*, at approximate cost	2,137.69	1,909.85
Inventory of Monograph on Acceptance Sampling, at cost	129.24	233.93
Inventory of Emblems, at cost	415.50	463.50
Furniture and fixtures, at cost less depreciation	2,088.78	2,192.42
Deferred Charges:		
Deferred Membership Directory expense		831.86
Other	945.55	1,007.87
	$64,030.95	$50,342.15

Liabilities and Net Worth

	1953	1952
Accounts payable	$10,430.60	$ 5,930.38
Deferred income (collections applicable to subsequent years)		
Dues	$16,827.00	$16,179.00
Subscriptions	5,817.77	5,524.48
Other	466.84	2,865.16
	$23,111.61	$24,568.64
Net Worth:		
Life Membership reserve	$ 3,579.92	$ 2,877.79
Surplus, per statement	26,908.82	16,965.34
	$30,488.74	$19,843.13
	$64,030.95	$50,342.15

American Statistical Association
Statement of Income and Surplus Accounts

	Year ended December 31, 1953	1952
Income:		
Dues—Current year	$38,607.00	$37,101.00
—Prior year	852.00	184.00
Life membership income	(90.13)	166.31
Subscriptions—*Journal*	10,134.80	9,543.00
—*American Statistician*	443.68	428.35
Advertising—*Journal*	1,415.99	1,376.75
—*American Statistician*	263.97	217.72
Sales—*Journal*	1,937.13	1,180.84
—*American Statistician*	148.02	82.65
—Acceptance Sampling	302.05	243.00
—Emblems, less cost of sales	83.28	11.00
—Membership Directory	22.50	45.00
—*Biometrics*	568.37	453.25
—Other	45.00	35.75
Mailing list income	911.11	704.17
Interest income	911.84	541.37
Annual meeting—see note	1,559.52	
Reimbursement of overhead expenses:		
Bureau of Mines Project	2,207.06	2,000.00
Miscellaneous	54.12	67.33
	$60,377.31	$54,381.49
Expense:		
Salaries	$12,260.07	$14,716.81
Publications—Schedule I	24,968.23	20,417.77
Promotion	861.89	702.30
Rent	2,400.00	2,400.00
Travel and secretarial expense	800.83	1,426.17
Supplies	2,240.92	2,434.54
Postage	1,849.26	1,163.45
Telephone and telegraph	729.02	563.71
Accounting services	970.00	970.00
Committee expense	1,269.75	575.50
Annual meeting expense	645.99	550.72
Miscellaneous expenses—Schedule I	1,437.87	1,928.83
	$50,433.83	$47,849.80
Excess of income over expense for the year	$ 9,943.48	$ 6,531.69
Add: Surplus account, at beginning of year	16,965.34	10,433.65
Surplus account, at end of year	$26,908.82	$16,965.34

Note: Includes $303.29 relating to receipts from 1952 meeting.

Schedule I

AMERICAN STATISTICAL ASSOCIATION

	Year ended December 31, 1953	1952
Publications:		
Journal—Printing	$15,118.77	$12,522.59
—Abstracts	750.00	
—Editorial expense	1,215.30	335.98
—Cost of old *Journals*	188.67	101.70
—Delivery charges	36.11	32.14
—Storage charges	114.00	96.00
	$17,422.85	$13,088.41
American Statistician	6,584.28	5,810.75
Acceptance Sampling	129.24	218.61
Membership Directory	831.86	1,300.00
	$24,968.23	$20,417.77
Miscellaneous Expense:		
Depreciation	$ 617.64	$ 555.87
Dues to other organizations	123.25	123.25
Bank charges	4.50	7.84
Taxes	314.89	559.41
Repairs and maintenance	153.25	206.95
Workmen's compensation insurance	24.53	26.91
Other	199.81	448.60
	$ 1,437.87	$ 1,928.83

SUMMARIES OF PAPERS DELIVERED AT THE 113th ANNUAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION IN WASHINGTON, D. C., DECEMBER 27 TO 30, 1953.

Edited by ARMEN A. ALCHIAN, *University of California (Los Angeles)*

The present section contains all available abstracts of papers presented at the 1953 national meeting of the American Statistical Association in Washington, D. C. The sequence of presentation here conforms to a grouping of abstracts according to the various sessions at which they were delivered.

PAPERS SUMMARIZED

Scientific and Professional Manpower: the Value and Limitations of a Statistical Approach. ELI GINZBERG, *Columbia University.*

A major shortcoming of American social science is its infatuation with the doctrine that if one only had a sufficient number of facts, one could solve any problem. This particular preconception is grounded in the following: (1) Our pragmatic position with its anti-theoretical bias. (2) The availability of funds to support expensive data collection undertakings. (3) Our desire to make progress quickly and our naive belief that the activities connected with data collecting are proof of progress. (4) Our belief that the "facts" will be able to resolve all difficulties and that there are no underlying value conflicts.

During the past few years much energy has been poorly directed because of an error in strategy: We started to collect large bodies of data without knowing what to do with them, and without having any clear idea of the key questions they were supposed to answer. Much of this data collection has been spearheaded by interested pressure groups to support a particular point of view. Much of the governmental and academic effort in the collection of data has been of dubious value. One of the most serious results of a preoccupation with statistical tabulations has been the blithe assumption that the concept of shortage (or balance) in professional manpower is a simple arithmetic relation between supply and demand. Similarly, there has been no proper attention paid to the role of substitution in matters of supply. Among the other significant facets of the problem that are not illuminated easily via statistics is that of utilization, which can of course greatly influence whether any given supply proves to be adequate or not. Entirely too little emphasis has been placed upon the fact that any particular manpower problem can probably be resolved in any one of a series of ways depending in large part upon the criteria that are employed. Perhaps the most serious shortcoming of all growing out of a quantitative approach is the inherent tendency contained therein to gloss over qualitative differences among professional persons and to see the problem primarily as one of numbers.

There are at least five major areas in which organized statistical efforts can make a significant contribution to the illumination of scientific and professional manpower problems: (1) By providing knowledge of the occupational structure; (2) By facilitating studies of the probable size of future supply; (3) Through contributing to the organized study of future demand by studying the strength of particular factors that have influenced demand in the previous time periods; (4) By helping to set out in a systematic fashion the incentive factors in different occupations which influence training and distribution of trained persons; and (5) by studies of the flow of individuals in and out of different types of employment, the strategically important question of convertibility of trained manpower can be illuminated. The rate at which a statistical approach can contribute along the foregoing lines to the understanding and solution of professional manpower problems will depend very greatly on the extent to which existing theory can be improved with respect to the concept of balance, the potentialities and limitations of "convertibility" in the use of highly trained persons, the elaboration of the significant relations between qualitative and quantitative considerations, and the factors determining different utilization levels.

Recent Advances in Statistics on Scientific and Professional Personnel. HAROLD GOLDSTEIN, *Bureau of Labor Statistics.*

Widespread interest in professional personnel has focused attention on statistical data in this field. Recent improvements include (a) improvement in the frequency, currency, detail, and over-all quality

of the data on output of trained personnel from the educational institutions; (b) development of techniques for registering members of professions; (c) active participation by professional societies in studies of their fields; (d) a beginning in the study of occupational mobility in the professions. The kinds of data we have and the gaps in information reflect the emphasis on large-scale statistical surveys, more intensive studies are needed to provide missing data.

Among the areas of work which should be productive are: I. Development of measures of qualitative differences among individuals to temper the purely quantitative information now available. II. Development of surveys of professionals via employing establishments, as an essential supplement to census-type and professional society membership type surveys. III. Investigation of factors and concepts underlying demand for professional personnel. The study of professional personnel within the context of general personnel and manpower problems is essential.

International Criminal Statistics. Bengt Helger.

In spite of repeated efforts over a hundred years to achieve greater uniformity of statistics on crime, a direct comparison of national series showing crime rates, and the trend of these rates, is scarcely possible. Criminal statistics cannot be taken at their face value, and international comparisons are particularly falacious without sufficient knowledge of all particulars, pertaining to the penal and judiciary system of countries, that are essential for an interpretation of national figures on crime. A new approach to the problem of international comparisons would consist in selecting for this purpose certain actions that are universally recognized as criminal and are of such a nature that they regularly become known to the authorities, and redefining these actions so as to give a uniform content to the series that are compared. Such data, stemming from the police or the investigating authorities, would be valuable as indicators of the frequency of crime, whereas court statistics, supplemented by information on suspended sentences, probation and parole, etc., have their value for studies of the treatment of offenders. Current statistics on crime can scarcely reveal the causes of criminality but may serve as a frame for further research into this subject by giving a general picture of the dimensions of criminality, the incidence of various types of criminality within socio-economic groups, and its evalution under the impact of social changes. In this connection, international comparisons present a great deal of interest.

The Study and Exploitation of Response Regions. G. E. P. Box and J. S. Hunter.

Techniques have been proposed by Box and Wilson for the study of response surfaces. Having reached a near-stationary region in the factor space, it is of extreme practical importance to study the local features of the response surface and to exploit the surface to best advantage. Designs are needed which will allow the fitting of a polynomial of degree d as efficiently and economically as possible. Optimal designs of first order (for the fitting of plane surfaces) have been described by Box elsewhere. The present paper discusses the problem of deriving optimal designs of second and higher order. It is shown that the principle of obtaining designs giving smallest variances for the estimates does not in the case of second and higher order designs result in a unique or a necessarily desirable solution of the problem. Some other principle seems to be required. The concept of information distributions is introduced. This distribution yields the information per observation obtained at a point on the fitted surface at any location in the factor space.

It is suggested that a *spherical* information distribution is desirable (i.e., a distribution such that the information is constant on spheres centered at the origin of the design). Designs which give a spherical information distribution satisfy the criteria that the variance-covariance matrix and the moment matrix of the design are invariant when the design is rotated. These new designs are therefore called "rotatable" designs. It is shown that the principle of rotability leads to optimal designs of first order. Second order rotatable designs are then derived. The designs are also considered from the point of view of the biases in the estimates which may be present if the assumed model is inadequate. Some confounding systems are discussed.

Developments in Statistical Techniques of the Last Decade of Special Interest to Economists: Sequential Analysis. Kenneth J. Arrow, *Stanford University.*

This paper is a survey of the fundamental ideas of sequential analysis and its implications for economics. The general conclusion is that sequential analysis as a statistical technique is likely to find few applications to econometrics because of the special problems of sampling in the typical problems of economic statistics, but that the type of reasoning underlying sequential analysis is likely to be of considerable importance in the theory of dynamic economics, particularly where actions with random consequences are under consideration. The technique of sequential analysis of statistical data is applicable when the experimenter can control the taking of additional observations on the basis of the results of earlier ones. The additional observations are possible at a cost. The technique is therefore not applicable to time series already collected, data collected as a by-product of other activities, to surveys

where it may be difficult to enumerate after each observation, or to multi-purpose surveys, where the stopping-point may be different for different questions asked. A brief development of the theory of sequential decision between two hypotheses is given from a Bayesian viewpoint. After a number of observations, the original *a priori* probabilities are transformed into a set of *a posteriori* probabilities, but the situation is essentially unchanged, except for the magnitudes of the probabilities. This simple observation enables an immediate derivation of the sequential probability-ratio test. A problem of choice among inventory policies is then stated. It is shown that the same type of reasoning used in the analysis of the sequential decision problem leads to a functional equation which can be solved to yield an optimal solution to the inventory problem.

The Problem of Non-normality, and Nonparametric Tests. WILLIAM H. KRUSKAL, *University of Chicago*.

Most statistical procedures used in practice are based upon the assumption of normality. The arguments usually given in defense of this assumption are stated and critically discussed. A method of circumventing the assumption of normality is the use of nonparametric statistical procedures. The advantages and disadvantages of these procedures are discussed in general, and some examples are given. The procedures of applied nonparametric analysis are classified, and key references are given.

A Cyclical Index of Highway Ton Miles. JAMES P. GEORGE.

This investigation has as its objective the construction of a composite weighted cyclical index of ton miles operated by trucks and truck combinations on all main and local rural roads in the United States. The index computes for any particular month the cyclical position of highway ton miles with respect to the computed normal or expected value as of that month. Polynomial curves of appropriate degree were fitted to the cyclical-irregular movement of highway ton miles over a 78-month period. Criteria, affording a test for goodness of fit, include the variance ratio or Snedecor's "F," the standard deviation of the residual variance, and the appearance of the plotted curve. This fitting (or smoothings process eliminates the irregular movements and causes those fluctuations attributable to cyclical force) to stand out in bold relief. It should be pointed out that the cyclical irregular movement to which the curve is fitted is in terms of composite weighted standard deviation units from "normal."

In consequence of this investigation the writer has been able to isolate and to delineate a structural pattern of the cyclical movement of highway ton miles which, with varying intensity, tends to repeat itself at approximately 32-month intervals. This mathematical function (the orthogonal polynomial curve) appears to postulate an underlying law, governing the behavior of highway ton miles. The structural pattern of the cyclical movement of highway ton miles is predicated upon an assumption of economic rhythm in the ton mile series. Even though trend and/or seasonal do change, the structural pattern of highway ton miles can be expected to continue, showing much the same contour as before. In other words, the contour of the structural pattern will tend to remain more or less constant irrespective of the level from which measured. From the standpoint of forecasting this is a most important consideration.

This investigation has been in part analytic and in part synthetic. In other words, it first concerned itself with a breaking down of the highway ton mile series into its component elements. In the latter stage, the investigation directed its efforts to a recombining of the constituent elements into a theoretical ton mile series. The synthesis makes possible an extrapolation or projection into the future It also serves to verify or confirm the correctness of the analysis by reconstructing the highway ton mile series from its constituents.

Approximate Tests for Comparisons of Rank Correlations. HERMAN O. HARTLEY.

Of the numerous measures of rank correlation the two best known are: Spearman's rank correlation r_s and Kendall's rank correlation t_k. Some results on the exact distribution of these measures are known for the case of 0-correlation and some moments for the case of ranks generated by a bivariate normal population with correlation coefficient ρ. In the latter case the present note investigates the z-transformation or r_s and t_k, both by the method of statistical differentials and by Monte Carlo calculation. It is found that for moderately large sample size both z-tranforms are approximately normal with variances approximately independent of ρ. This property results in simple tests of significance for comparisons of rank correlations so transformed. The restriction to ranks generated by samples from bivariate normal samples is lifted and the tests shown to apply to a much wider class of ranks.

Optimum Cluster Size. HOWARD L. JONES.

When a two-stage sampling procedure consists in selecting a number of clusters of equal size, and the total cost is a linear function of the number of clusters and the number of individuals selected, the optimum cluster size is a simple function of the parameters of the cost function and the intraclass and interclass correlation coefficients. Suggestions for estimating these parameters and coefficients are

proposed and discussed for an illustrative example. The more general situation is then examined where the number of attempted selections is the same for every cluster, but the actual number of individuals inspected varies from cluster to cluster.

A Note on the Use of Normal Probability Paper. HERMAN CHERNOFF and GERALD J. LIEBERMAN.

This paper illustrates, with a special example, that the graphical technique to be applied to a problem should depend to a large extent on the use to which the graph is to be put. In particular, we treat the problem of selecting the representation of a sample on normal probability paper when it is desired to obtain "optimum graphical" estimates of the mean ξ and standard deviation σ of a normal distribution.

Two Nonparametric Tests Using the Method of Ranks for Testing the Randomness of Samples Drawn from Finite Populations. MITCHELL O. LOCKS, *University of Oklahoma*.

In this paper are developed two nonparametric tests based on the method of ranks, both of which may be used to test the randomness (or representativeness) of samples drawn from finite populations. The theory behind the development of these tests is essentially the same as that governing tests for randomness which use the original values. Since the application of these tests requires the ranking of all of the items in the population according to some numerical characteristic, the tests can be applied only if values with respect to at least one numerical characteristic are available for every single item in the population. Admittedly, this restricts the scope of usefulness of these tests. However, it is believed that the tests can be used more profitably for certain types of problems (e.g. tests for randomness of samples drawn from J-shaped and U-shaped populations) than other statistical tests for randomness now in use.

A Two Sample Procedure for Linear Discrimination in Normal Samples. JACK MOSHMAN, *Oak Ridge National Laboratory*.

The problem of discrimination from a normal linear regression model considered is that of obtaining a confidence interval for x corresponding to an observed y when the expectation of y, given x is $E(y\,x)=\alpha^*+\beta^*x$. Various incongruities are discussed which result from a naive approach. A deterministic one sample procedure is shown to have two difficulties: (1) The procedure is inefficient in that the confidence interval obtained is larger than a specified $1-\gamma$; (2) The excess is a function of the unknown variance. A two sample procedure is exhibited which makes the excess over $1-\gamma$ to be less than a specified $\delta(0<\delta<\gamma)$ whenever $|\beta^*|<l$ for any predetermined $l>0$ and the excess is independent of the variance.

Probability of Acceptance for Sampling Plans Based on Average-Standard Deviation Acceptance Criterion. HARRY G. ROMIG.

The evaluation of a variables sampling plan using Average-Standard Deviation Acceptance Criterion requires Probability of Acceptance values for determining its Operating Characteristic. The mathematical relations are presented for computing such probabilities as developed in 1934 assuming a normal law distribution for the parent population and resultant exact distributions for averages and standard deviations where the correlation between averages and standard deviations is zero for two cases: 1. Sample size n large: distributions of averages and standard deviations assumed normal; and 2. Sample size n small: distribution of averages normal and of standard deviations non-normal. These results may be compared with later work started in 1947 at Stanford University by Goode, Bowker, Ireson, and Resnikoff.

For the two cases a cutting plane slices the frequency surface for averages and standard deviations and volumes under the surface are determined for shifts in level (average), variability (standard deviation), or both, due to changes in the system of causes. Levels of incoming quality with respect to engineering requirements and the protection desired for any sampling plan determine the location of the cutting planes used for evaluation. The Average-Standard Deviation Acceptance Criterion is set up for either a minimum or maximum engineering limit for sample values of averages and standard deviations. For such criteria, relations are provided for determining the volume under the surface to the right or left of the cutting plane covering both max. and min. limits. These relative volumes provide the desired probability of acceptance values for any postulated incoming quality p'. A simple approximation for n small is given. Exact and approximate P values for samples of 5 are presented for two sampling plans for evaluating the error of the approximation. A graphical solution of the multiple integrals is described. These probability of acceptance values must be obtained in order properly to set up various types of variables sampling plans.

Time Series Factor Analysis with an Economic Application. ERNEST W. BROWN and MAX A. WOODBURY

It has been recognized that the application of many statistical techniques to time series is invalid due to the correlation of successive observations. Among the techniques affected is factor analysis, since many of the underlying relations, which a factor analysis should uncover in a time series, act with a time delay, and hence would not be discovered. An appropriate definition of a factor and a technique of time series factor analysis is formulated and discussed and an application to a set of ten economic time series is made. The ten series cover national product, income and employment. In factor analysis, if there are n original variables, no more than about $n/2$ factors can be identified. An interesting feature of time series factor analysis is that this does not necessarily hold true. In fact, factor analysis of time series may require, and allow identification of, more factors than variables.

A Stochastic Model for the Selection of Macronuclear Units in Paramecium Growth. A. W. KIMBALL and A. S. HOUSEHOLDER, *Oak Ridge National Laboratory.*

Prior to division, the macronuclear units in *Paramecium* (see, for example, Sonneborn, *Ann. Rev Microb.*, pp. 55–80, 1949) presumably double in number in a manner similar to chromosome doubling at mitosis. During the process of division each daughter animal receives approximately one-half of the units present. Thus in any population of *Paramecia*, each animal has about the same number of units. If there are n units per animal, there may be as many as n different types of units, or all units may be alike, and any combination between these extremes is also possible. Such combinations are called states. Given the state of a single animal and given an hypothesis about the selection of macronuclear units by daughter cells, the probability distribution of states after N divisions may be computed readily by the methods of stochastic processes. Results have been obtained for the hypothesis of completely random selection and have been compared with experimental data.

Stochastic Processes and the Study of Growth Phenomena. A. T. REID, *Columbia University.*

This paper is divided into three parts: (1) Introduction, (2) Construction of stochastic models, and (3) Statistical inference in stochastic models. In Part 1 we discuss the deterministic and stochastic approaches to the study of growth phenomena; and give an introduction to the theory of branching stochastic processes as developed by Bellman and Harris. We consider in Part 2 the construction of various stochastic models for growth using the above theory. Models for birth, birth-and-death, and mutation processes are discussed. The use of these formal models in the study of epidemics and rumor spread, as well as in the study of bacterial growth, is pointed out. In Part 3 we discuss problems of estimation and testing associated with stochastic growth processes. Previous work is reviewed, and some recent investigations on sequential decision problems for branching processes is discussed.

The New Bureau of Labor Statistics Indexes of Productivity in Manufacturing. LEON GREENBERG and ALLAN SEARLE

The Bureau of Labor Statistics is planning to publish several series of productivity indexes in manufacturing covering the years 1939, and 1947 through 1952. Two of the series will show changes in man-hour requirements for physical production in manufacturing. A third series, based on concepts similar to the gross national product approach, will show changes in man-hour requirements for dollar value added in manufacturing and will reflect the influence of changing inputs of material, supplies and fuel. The various measures fill in a statistical gap of 13 years for which reliable all-manufacturing productivity statistics are not available.

The indexes are being derived from secondary sources of production and man-hour data. In relating these two statistics to derive productivity ratios numerous data problems are encountered. Not all of them can be solved immediately and the indexes released by the Bureau will not be precision instruments. But they will still be useful and necessary indicators in the broad areas of economic and business analysis, manpower planning and productive efficiency.

Problems and Prospects of Criminal Statistics in the United States. THORSTEN SELLIN.

Criminal statistics issued by agencies dealing with crime or criminals have been assumed to be important sources of data for the scientific study of the trends of criminality, the characteristics of offenders and the efficiency and effectiveness of the agencies dealing with them. The dearth of good criminal statistics in the United States is discussed and reasons for this situation considered. The main part of the paper examines the problem of how such statistics can be used in the study of trends of criminality and the characteristics of the offender, and raises a number of questions designed to stimulate a discussion of this problem. Illustrations are offered pointing to the need of improving certain official reports now issued, attention being drawn in particular to the need for some revision of the standard classification of offenses employed by many national and state agencies and the need for more carefully computed crime rates than those now being published.

Organization and Functions of a Complete Centralized Statistical Center at a Land Grant College. T. A. BANCROFT, *Iowa State College.*

The position is taken that there is no unique optimum organization or program for a statistical center which will work equally well at all universities or even at all land grant colleges. Further, even at the same institution either one of several alternative arrangements might work equally well. A *complete* statistical center should provide: (i) a research and teaching program in statistics per se in order to develop new statistical theory and methodology and train statisticians; (ii) a service teaching program to provide for basic general courses in theory and methods and specialized courses in statistics for students majoring at the undergraduate and graduate levels in some other substantive subject matter area; (iii) a consulting service program, i.e. recognized and budgeted time for various staff members of the statistical center to consult with research workers on investigations involving the use of statistical theory and methods; and (iv) a computing service for the programming and analyses of data resulting from research investigations. These four main objectives could be accomplished by a separate department of statistics, or, in case of a small group, a sub-department, and a campus-wide statistical laboratory of institute status.

Role of a Centralized Statistical Organization in a University. J. G. DARROCH, *Washington State College.*

The recommended relationship between the statistical organization and the various research units of the University may be set forth as it affects some of the major units. (1) The Agricultural Experiment Station—two or more staff members assigned to service this unit, with the privilege of calling on their associates to aid in special problems or to share the load at seasonal peaks. The consultor should be permitted to exhibit a preference among the possible consultants, if he so wishes. Some Directors might prefer to have the consultants attached to their staff but the preference would seem to be to house them with the statistical group, unless there is a distance problem, and have them administratively responsible to the head of this same group. (2) The Department of Mathematics—the liaison here, established by joint appointment or some other effective means, needs to be rather close. The relationship here on consulting problems may well be one of the statistician consulting with the mathematician. (3) The Social Science Departments—the needs here could best be served by staff members specifically trained in certain statistical disciplines, again a matter of assignment. (4) Engineering, Medicine and other fields—while specialized knowledge would be an advantage there are many problems where a straightforward statistical approach would be of material assistance to the research program. The inference here is that much can be done by the general consulting staff in the early development of a consulting load. As the loads from these areas build up it would be advisable to add personnel with background training in the various fields. (5) The computing facility—for statistical problems the consultant should provide the liaison between the worker and this unit. Thus it would appear desirable that the computing unit be an integral part of the organization under discussion.

The other aspect of the consultation program that should be mentioned is that having to do with the needs of graduate students. The most desirable pattern would seem to be that of the graduate student and his major professor both participating at least in the initial consultation. At that time the relationship of the consultant to the research problem should be fairly clear, and responsibility for direction of statistical matters and for computing decided upon.

The Role of a Centralized Statistical Organization in a College with Respect to the Reviewing of Manuscripts and the Promotion of Statistical Research. H. C. FRYER.

Statistical methods should not be forced on the research staff of the college or university, but the arguments for their use should be brought to the attention of the administrators of the research programs. To that end the following system is recommended: (a) Every proposed new project should be studied by a committee including a representative of each department involved, the statistical group being considered to be involved whenever the proposed experiment is to produce sampling data. (b) This committee is to make written recommendations to the directors of the research programs, including the presentation of minority recommendations when decisions are not unanimous. (c) When manuscripts result from the research they are subject to the rules of (a) and (b).

Most colleges and universities have limited funds, and someone must be convinced that the research will produce the greatest good for the money expended. The consumable end-product is some form of applied statistics; hence, it is suggested that every statistician in the organization should divide his or her time among teaching, statistical consultation with those doing applied research, and pure statistical research. The division should not necessarily be in thirds; emphasis should be on the activity best adapted to each person's interests and abilities. This plan should point statistical research toward problems the applied scientists need to have solved, and should keep the statisticians from getting into intellectual "ruts." Some may think the cost of paying several persons capable of research instead of one or two who do nothing else will be greater. Many of our young Ph.D.'s are capable of research,

and would gain valuable experience in such positions. If they like this multi-purpose activity and choose to stay with it, their productivity should increase with the cost of keeping them indefinitely. Before they settle permanently on such a job they will realize there is a limit to the pay they can receive; and they can go elsewhere if they prefer. In the meantime the school will benefit by the newness and the vigor of their statistical activities. Those schools in which money is no limitation and which do not have to worry about the applications of their research products may wish to supplement the above plan with some eminent research personnel who wish to do full-time theoretical research.

Staffing a Central Statistical Organization. HERBERT A. MEYER.

The number and qualifications of the members of the staff of the Central Statistical Organization depend upon the functions and responsibilities that are assigned to the Central Organization. These might be grouped under the following heads: (1) Consultation, (2) Teaching—theory and application of statistical methods, (3) Evaluation of manuscripts where statistical methods are used, (4) Research, theoretical and applied, and (5) Computing Service (machine room).

In a large university, the technical staff of the Central Statistical Organization should be composed of specialists in various statistical fields. Where teaching is one of the functions, the staff will have to be quite large and those who teach should also do research and consultation. Where teaching is decentralized, joint appointments of at least some of the teachers would seem to be desirable in order to bring the Central Organization in close relationship with those departments where problems in the application of statistical methods generally arise. The plan of joint appointments is working quite satisfactorily in the University of Florida, where the teaching of statistics is not centralized.

In a small university, the same general organization would seem desirable, but if funds are limited and demands less varied, the staff will be small. This calls for greater versatility on the part of the members of the staff. While in a large university the Central Statistical Organization may be, and generally is, an independent unit, in a smaller institution it may be necessary for financial or other reasons to make it a branch of the mathematics or some other department. The computation service, or machine room, is an important section of the Central Statistical Organization. To be effective, it must be staffed with a competent supervisor familiar with statistical methods, as well as with competent keypunch operators and machine operators. Where courses are given in the university in the operation of IBM and other types of computing machines, these can serve to train potential replacements and additions to the card machine computing section. It is a difficult problem to find persons with the desired qualifications who are willing to accept the starting university salaries because of the demand in industry and government for statisticians at salaries above those in most universities; and this, the problem of finding a staff at all, is perhaps the most difficult of all staffing problems.

Responsibility of a Centralized Statistical Organization in a University for Training in Statistics. W. ALLEN WALLIS.

Departments in fields other than statistics normally do not have faculties qualified to teach current statistical methods, so the responsibility for statistics courses should not rest with them. On the other hand, statistics departments vested with sole authority over the statistics courses may tend to overemphasize statistics and its mathematical prerequisites at the expense of the students' subject matter studies. At Chicago, we have developed what looks to be a promising method for reducing duplication of statistics courses in the various departments and raising the level of instruction, while avoiding the pitfalls of authority centered in the statistics department.

Our first step was to introduce a new elementary course taught by one of our own men. After this course was well developed, people teaching statistics in various departments were invited to merge their course with ours. In general, their reaction was to introduce our syllabus and other materials but retain their instructors and the independence of their courses. The syllabus and teaching materials were revised somewhat in the light of suggestions and criticisms from these instructors in other departments, and then one or two departments did merge their courses with ours. As those departments reported favorably on their experience, others took the plunge. Essentially the same evolution has occurred with the advanced course on statistical inference.

With this system, each department lists the unified course in its own section of the catalog with its own number; but all designate the same room and instructor. Any department could secede simply by making a different room assignment in its Announcement. Actually, this would not be likely to happen without consultation with the statistics department, since other departments know that we are sensitive to their needs and limitations as well as to our own standards and requirements. Unification of statistics courses prevents their being too closely tied to a narrow range of substantive problems, encourages the use of examples from a variety of fields, and is generally an advantage pedagogically.

Whether the device used at Chicago would work at other universities is a question that must be answered in the context of each institution. If, with the elimination of a course to reduce duplication,

a department suffers cuts in appointments and budget, then it will be less ready to reduce the duplication. Some budgetary readjustments may be in order; but so long as cuts are not allowed to work to the direct disadvantage of those who help achieve the gains, many of the obstacles to eliminating duplication in universities seem to be avoided.

Trends and Cycles in German Wages. GERHARD BRY.

The paper compares German, English, and American long-term trends in wage levels and in wage structure, as well as their cyclical behavior, from 1871 to 1953. Hourly money wages are shown to have increased about seven fold in Great Britain, eightfold in Germany, and about ten fold in the United States. The divergence in the increase of real wages—two and one-half fold for Germany and Great Britain and fourfold for the United States—indicates the close relationship between wages and economic and political fortunes in these countries. A tendency toward decreasing differentials can be observed in the three countries. This trend towards greater equality is relatively clear in German skill, sex, age, and regional differentials. It is suggested also in city size and industrial differentials. Basically, this trend follows from the industrialization process itself, which leads to reduced differences between wage earners.

Although wage rates respond to general business contractions—by decreasing rates of growth, stagnation, or declines—there is disclosed a definite downward rigidity of wage rates. During the period 1871–1953 only two substantial declines in German money wage rates occurred. Earnings have larger cyclical amplitudes than rates because overtime, output bonuses, and hours worked are cyclically more sensitive. Even in the Great Depression, real weekly earnings in Germany and in the United States declined by only 15 per cent. In both countries the real sufferers from the Great Depression were the unemployed rather than the employed workers.

Union Impact on Wage Structures. H. M. DOUTY, *Bureau of Labor Statistics.*

A series of studies by the Bureau of Labor Statistics indicate that relative wage differentials among occupations have declined markedly in American industry, particularly during recent years. Over the past half century, several factors calculated to narrow job differentials in the long run can be distinguished. Among these are the sharp decline in immigration during World War I and the subsequent adoption of a restrictive immigration policy; a declining birth rate until the 1940's; a rising level of education and training among the working population; and the mechanization of large areas of unskilled work. These broad labor market forces clearly produced some tendency for relative job differentials to diminish. The decline in job differentials during the past decade, however, was much sharper than would have been anticipated on the basis of long-run labor market forces alone. One factor was governmental wage policy, notably in the economic stabilization program during World War II. Of major importance was a marked tendency for unions, beginning in the defense period in 1941, to formulate their wage demands in terms of uniform money increases (and hence unequal percentage increases among jobs) and settlements to be made in this fashion.

Among the reasons for this development were: (1) uniform money increases tend to appear more equitable in an inflationary period when wage increases are designed largely to offset increases in living costs; (2) it is politically easier, especially in an inflationary situation, for union leadership to press for uniform money increases; (3) within limits, skilled workers may be content with the maintenance of absolute wage differentials, even though their relative wage position is deteriorating; (4) on the whole, employers during the past decade have seemed more concerned with the size of negotiated wage increases than with the form of their distribution; (5) the structure of wages probably seems less important than the general level of wages, at least in inflationary periods. The decline in relative wage differentials among jobs has been much more marked in some industries than in others. Divergence in experience within manufacturing is well-illustrated in the basic steel and automobile industries. Assuming reasonable economic stability, it is probable that relative job differentials will receive considerable attention during the next few years from unions and employers. Special wage adjustments for skilled workers in some industries during 1953 point in this direction. Systematic review of the structure of job rates may occur in a significant number of situations.

Wages Since 1914. LEO WOLMAN, *Columbia University.*

This paper deals with one segment of wages in this country, the behavior of hourly earnings in five industries—manufacturing, class I railroads, building, anthracite and bituminous coal mining. Average hourly earnings, gross and net, are presented as an approximation of changes in the price of labor during these 40 years. Despite interruptions in their upward movement, money and real hourly wages multiplied many times over the period. All five groups shared in the rise, but unequally. Anthracite coal topped the list with nearly a 10-fold increase; building trades' rates rose least, less than 6-fold. Real hourly wages likewise mounted, but, of course, less than money wages. Anthracite real wages rose

3.5 times, manufacturing 2.8, and building over 2 times. The interesting question is, what economic conditions accounted for this multiplication in real and money wages? Using manufacturing wages as an illustration, it is clear that the bulk of the total money advance of $1.47 an hour was a product of the economic conditions of war and postwar boom, since only 10 cents of this amount were added between 1920 and 1940.

When money wages are converted into real wages, surprising results are obtained. For real wages showed marked improvement without apparent regard to movements in money wages, business conditions or union organization. Comparing the course of money and real wages in the two wars one encounters many difficulties. The first was shorter and didn't have wage controls. In several industries wages in World War II were fixed in long-term contracts. Bearing these factors in mind, the available evidence suggests that money and real wages rose more between 1916 and 1918 than from 1941 to 1945. This record of 40 years suggests that the period beginning in 1914 can be described as a high-wage era in the sense that the increases of this period were not matched by the wage advances of the preceding half-century. On the question of organized labor's effect on wages, the evidence is conflicting. Attention should be called, however, to the striking rise in manufacturing real wages between 1929 and 1940 when average unemployment stood at a consistently high level and when the only tenable explanation of the rise in real wages must be public policy and trade unionism.

Limitations of Consumer Credit Statistics. Ernst A. Dauer, *Household Finance Corporation.*

Definition and content. All elements in the Federal Reserve Board's definition are relatively simple and appear clear. However, differences arise in attempting to apply the definition, because there are legitimate differences in point of view, and because different people expect the figures to serve different purposes. Even the Federal Reserve Board does not seem to appreciate fully that difficulty also arises because the consumer credit total figure is used, as a rule, for an altogether different type of analysis than the break-downs. Their decisions on the items to be included for one type of analysis make it impossible for the resultant figures to serve properly in another analysis. The basic published estimates are limited to the amounts of credit outstanding. Figures showing the flow of consumer credit, if added, would provide a much more complete understanding of current developments.

Methods of estimation. The recent revision has distinctly improved the technical accuracy and coverage of the estimates. Even with their shortcomings, they are substantially better than the statistics available in most economic areas. I am concerned about the broader, non-technical aspects of the problem. Trained statisticians are aware: that *all economic* statistics measure differences of degree; that all classification is subjective; and that every problem really requires its own set of statistics, if they are to be truly germane. Most users, lacking formal training, are misled by the aura of accuracy and precision which surrounds the mere issuance of estimates by a federal government agency. A clear, concise, and simple disclaimer is necessary to dispel the mirage of infallibility.

On the surface the estimates provided for commercial banks—the most important single type of holder—appear to be the most thoroughly grounded of any of the series. Yet, the whole procedure involves guesswork, because the banker frequently does not know the use of each loan, at the time it is made; and experience shows that many respondents give whatever figures are easiest when condition reports are prepared, semi-annually. The correction factor applied by the Board to exclude the non-consumer portion from the reported figures is based upon a single survey—a sample which seems wholly inadequate. The validity of using the findings of this single survey from 1939 to the present time, and into the indefinite future, is questionable.

Presentation and classification. The presentation of data in the Federal Reserve *Bulletin* represents a distinct improvement over the old tables. This is true, particularly, of the functional break-down within the instalment credit sector. Other basic tables break down the amount of instalment and non-instalment credit respectively, according to holder. Here the decision to treat sales finance companies on a consolidated basis (including the operations of cash lending subsidiaries) appears unwise and should be revised. To attain the stated objective of the Board (to present the data in such detail that it can be taken apart and put together by analysts with various interests and various types of problems), much finer break-downs are necessary—and could be obtained if the Board made a serious attempt to do so.

Recent Revisions on Consumer Credit Statistics. Homer Jones, *Federal Reserve Board.*

The Federal Reserve Board in 1953 completed a revision of the statistics regarding consumer credit outstanding from 1939 to date. The new data are the first thorough revision of the series since their inception in 1940 and were made possible by the results of the 1948 Census of Business. The revision resulted in a substantial increase of estimates of consumer instalment credit outstanding and a substantial reduction in estimates of charge accounts outstanding and a moderate increase in total consumer credit. The revised data and a description of the process of revision are presented in the Federal

Reserve *Bulletin* for April 1953. Technical aspects of the revision are discussed in a supplementary pamphlet issued by the Board. While the revision of the series was desirable and necessary, it appears that the unrevised series had never been misleading as to order of magnitude of amount outstanding or of change in outstandings. Furthermore, it is believed that the use of annual Census Bureau data from the retail trade survey will provide a much more accurate basis for estimates in the future.

In connection with the revision of these series, a thorough evaluation has been made of the concepts involved and the means of making current estimates. Denomination of the series as "short- and intermediate-term consumer credit" has emphasized the fact that it covers only part of consumer credit, since house mortgages owed by owner-occupants are excluded. The "service credit" series has been thoroughly revised, primarily by basing its major component, amounts owed to medical practitioners, upon data from the Survey of Consumer Finances. Treatment of data supplied by commercial banks has been improved upon a basis of the results of sample surveys.

The Function of the Outside Consultative Committee in the Revision of Governmental Statistics. B. D. MUDGETT.

In the revision of governmental statistics, there are five parties at interest—employers, employees, the government bureau in question, the outside public, and the whole body of statisticians interested in the particular statistical output. Employers' and employees' interests may or may not coincide with the ends sought in a revision; advantage accrues to the general public from a strong and healthy economy and statistical procedures that assist this long run objective are in the interest of the general public. The bureau personnel is in general highly competent but is beset both by routine operations that resist change and by outside pressures of groups dominated by their self interest. The whole body of scientists, as scientists, must be presumed objective and their pressures therefore lead in the direction of constant improvement of statistical processes. Herein lies the good earth that must be worked by the committee of "experts" when they are asked to sit with bureau personnel in the revision of important bodies of governmental statistics. It is this situation that gives them a function the performance of which may aid in the improvement of government statistics. They operate to perform three tasks: (1) Better theory—the clarification of concepts back of particular bodies of data or particular statistical tools through give and take discussion between consultants and bureau personnel; (2) better liaison between government bureaus and the outside public, leading to education of the public on the significance of the work of bureaus and on the competence of staff; and (3) encouragement toward working conditions that will attract competent personnel and lead to improved career service in governmental agencies

A Follow-up Study of Mortality in World War II Prisoners of War. BERNARD M. COHEN, *National Research Council*, and MAURICE Z. COOPER, *Veterans Administration.*

Using tested methods of follow-up by matching existing military, Veterans Administration, and other records, and by questionnaire survey, the Committee on Veterans Medical Problems, National Research Council, is conducting for the VA a follow-up study of mortality, morbidity, disability, and adjustment in U. S. white male Army personnel who were prisoners of war in World War II and liberated alive. Representative samples of both Pacific and European ex-prisoners, together with appropriate control groups, are being followed. This initial report is confined to mortality as observed in the first six years after liberation, and deals primarily with methods. The main methodological features are: (1) Demonstration from previous experience of the virtual completeness of VA death records of servicemen and veterans, permitting accurate measurement of mortality differentials, and providing reliable estimates of possible error where the differentials are small. (2) Selection of control samples from Army unit records to match controls with prisoners in proportion of ground and air forces, officers and enlisted men, and time distribution of and opportunity for capture, the latter two factors being better approximated in the European than in the Pacific area. (3) Use of (a) an additional control, a sample of veterans generally, to evaluate the effects on mortality of selection for military service, and (b) population life table mortalities as an intermediate device to derive stable, annual, age-specific mortality expectations.

The main findings are: (1) A marked excess of mortality in the Pacific prisoners, somewhat concentrated in the first two years after liberation, and no excess of mortality in the European prisoners. (2) In the mortality excess of Pacific prisoners, tuberculosis and automobile accidents are the most conspicuous causes of death.

Insurance Mortality Investigations of Physical Impairments. E. A. LEW, *Metropolitan Life Insurance Company.*

Life insurance mortality investigations may be regarded as a classical example of long range follow-up studies. They have usually dealt with relatively large numbers of persons who have been automatically traced for long periods of time through the circumstance of their being insured. The pro-

cedures for such studies were fully developed by actuaries more than a hundred years ago. The paper draws attention to the salient features of several types of life insurance mortality investigations and to the essential procedures used in such long range follow-up studies of physically impaired lives. It outlines the scope and the principal findings of the more important investigations of physical impairments, and discusses both the limitations and the special value of the data for public health and medicine.

The following mortality investigations of physical impairments are specifically referred to: Medico-Actuarial Mortality Investigations, 1912–1914; Medical Impairment Study, 1929 and its Supplement; Medical Impairment Study, 1936; Medical Impairment Study, 1938; Blood Pressure Study, 1939 and Its Supplement; several recent studies of the effect of build on mortality; several recent studies based on follow-ups of disabled policyholders.

Reference is also made to the Impairment Study, 1951, a comprehensive intercompany investigation of the past fifteen years' experience under some 132 groups of physical impairments, which is due to be published by the Society of Actuaries in the spring of 1954.

The findings of these investigations have made it possible to express the long range prognoses for a wide variety of impairments in numerical terms. They have been particularly valuable in shedding light on those impairments that fall in the broad range between good health and disease as recognized by clinicians, for example, overweight and moderate elevation in blood pressure. Follow-up studies of policyholders who recover after incurring a disability while insured have produced data indicative of the long range prognoses for the more serious impairments, such as advanced tuberculosis and coronary occlusion.

Factors in Interpreting Mortality After Retirement. ROBERT J. MYERS, *Social Security Administration.*

Currently there is considerable discussion as to the advantages of individuals continuing in employment beyond age 65 rather than being forced to retire compulsorily. Such advantages accrue both to the individual and to the nation. One of the subsidiary advantages frequently claimed is that an individual who is compelled to retire will lose his vitality and die much earlier than if he were allowed to continue in gainful employment. This runs contrary to the viewpoint frequently expressed several decades or more ago that workers were being kept in harness until they dropped dead from exhaustion rather than being allowed to spend their declining years in peace and leisure.

Unfortunately, specific and reliable data as to the effect of retirement on mortality are not available. The analysis is complicated by the question as to whether people retire because they are disabled and thus subject to high mortality or whether the retirement itself produces the high mortality. Clear evidence is available that retired persons do have higher mortality than active workers and the general population at the same ages, especially in the first few years after retirement. These data are presented for various Governmental retirement systems and for certain selected non-Governmental programs. Consideration is also given for various types of retirement systems as to how the resulting mortality experience may develop and what biases and limitations may be present solely because of the particular provisions of the plan rather than because of any underlying mortality effects.

Some Observations on the Inequality of Incomes. HERMAN P. MILLER

Throughout history philosophers have speculated about reasons for the inequality of incomes. Various explanations have been offered. Sone have stressed ability, others chance, and still others institutional factors which give the children of the wealthy an undue advantage. This paper offers a statistical explanation for the skewness of the income curve. The thesis is that the skewed income distribution reflects the merging of several symmetrical curves which differ only with respect to the level and spread of incomes. To explain income inequality it is first necessary to understand the reasons for the differences in the component parts of the income curve.

Much of the skewness of the income curve is due to the inclusion of women in the distribution The difference between income distributions for men and women has little to do with chance, ability, or the possession of private wealth. Among men, nearly three-fourths of the highest income group (over $10,000) are independent professionals, businessmen, or managers. To the extent that there is freedom of entry into these occupations, income differences between these groups and others may merely represent the payment by society for rare skills or risk taking. The facts regarding freedom of entry are not now adequately known.

Value of Dwellings in Relation to Income. MARGARET REID, *University of Chicago.*

This analysis of value of owner-occupied dwellings in relation to income using data from the 1950 census is still in a preliminary stage. Major findings include the following: (1) Coefficients of elasticity of value of dwelling in relation to income with grouping by income of primary families and individuals occupying them had a considerable range, e.g. from .18 in Cleveland to .63 in Birmingham. For S.M As (Standard Metropolitan Areas) in general it appears to be around .30. ($V=a+bI$, where V equals the

logarithm of value of owner-occupied dwelling and I equals the logarithm of income of primary families and individuals occupying them.) Many factors appear to contribute to this relatively low coefficient as well as to differences among S.M.A., e.g. distribution of hoseholds by age and sex of head and random variations in income. Households with head 65 years or older or with female head tend to have a high value-income ratio and tend also to concentrate at low incomes. In addition, random variations of incomes tend to concentrate at low incomes persons with a transitory decline in income and at high levels those with a transitory increase. This distribution of transitory incomes interferes seriously with differences in current income indicating what families tend to do when changes occur in income status. However the coefficient of .30 is quite similar to the corresponding coefficient of imputed rent of owner-occupied dwellings in relation to income, for places surveyed by the U. S. Department of Labor during the thirties. (2) The average value of owner-occupied dwellings for various S.M.As., standardized for income level of the family, is related to the income of the S.M.A. This is consistent with earlier observations that expenditure levels appear to be a function of the income of the community as well as the current income of families. There seems good reason to believe tha. if the effect of random variations in family income is eliminated the association of value of dwelling to community income as a separate factor is no longer present. (3) Coefficients of elasticity of value in relation to income from a grouping of households by value of dwelling tend to approximate 2.0. Since value of dwelling probably has less random variation than current income, it may be that this regression provides a better measure of the effect of difference in income status on the value of dwelling than does a grouping by current income Differences in level of the value curve among S.M.As. from this classification are related to age distribution and not to income of the community. (4) Intergroup value-income relations among major portions of the New York S.M.A. and among census tracts of several places yielded coefficients of elasticity of value of dwelling to income around 2.0.

Changing Geographic Patterns of Migration in the United States. Henry S Shryock, Jr., *U. S. Bureau of the Census.*

In order to study trends in net migration by States and possible changes in the overall pattern of interstate migration, figures were compiled from a variety of sources. Many of the data used were estimates especially prepared for this purpose. The following nine periods were included: 1950–1952, 1949–1950, 1945–1950, 1942–1945, 1940–1942, 1935–1940, 1930–1935, 1920–1930, and 1910–1920. Using the States as units, product-moment correlation coefficients were computed for all pairs of time periods Of these 36 correlations, 31 were significantly different from zero at the .05 level and were all positive. The median value of all the correlation coefficients was +.58, and the range was from −.03 to +.93. There did not appear to be any systematic changes in this pattern over time; adjacent periods were not more highly correlated than nonadjacent periods. Furthermore, the pattern by States did not seem to be related to the gross volume of interstate migration of a given period. There were indications that the pattern was sensitive to economic conditions and to war conditions. The relationship was not a simple one, however.

These conditions have more obvious effect on the figures for some of the individual States. Many of these States do show fairly definite trends in net migration. The year from April 1949 to April 1950 had the most atypical pattern of interstate migration. For this year and the period 1935–1940, statistics are available on migration in both directions between pairs of States Four States (California, Michigan, Tennessee, and Texas) were selected for intensive examination. For each of these States, the important interchanges with other States were listed, and changes between the two periods were noted. Some of the most important shifts in migratory currents were thus detected. The quality of the basic data, additional information needed, and some questions for future examination are also discussed.

The Problem of Improving Mineral Statistics. J. E. Morton, *Cornell University.*

The problem of improving mineral statistics is only a part of a universal problem: the improvement of the systematic mass production of economic data in general. That the collection of mineral statistics should create a special problem may be attributed to the late but rapidly growing interest in the mineral sector of our economy; in part, the problem is due to the technological peculiarities of the extractive industries. Weak-spots and other "pathologies" in a given fact finding machinery are easily identified if one projects such machinery against generally accepted requirements and standards. Examples of such requirements applying to any efficient and well organized mass data production process are: (1) The proper specification of the end product; (2) The efficient production of the data while adhering to accepted statistical standards; (3) The quality control of the product to protect the data-consumer and to yield a basis for continuing improvement of the data production process. Comparing—from the above points of view—mineral statistics with other major types of economic data collection systems, with agricultural, manufacturing, population and employment statistics, one must admit that mineral statistics frequently violate the above sketched requirements. The resulting weaknesses have been reflected in such diagnostic reports as that by the Hoover Commission, the President's Material Policy Commis-

sion and the American Statistical Association's Survey of the Statistical Operations of the Bureau of Mines.

In conclusion, it appears that several basic steps will have to be undertaken before special therapeutical measures can be applied effectively. These steps are: (1) The development of a broad, yet specific, mineral statistics policy which will not only recognize the difference between the needs of the technologist on one hand and those of the business analyst and economist on the other, but also the discrepancy between the commodity and the industry-wide point of view. (2) The adherence to modern statistical standards and the development of new methods and procedures where justified by the peculiarities of the extractive industries. (3) The reconsideration of the administrative framework within which efficient production of mineral statistics is to develop, including the very important question of how to generate, attract, and best utilize the particular kind of rare combination of talents and skills needed for the successful operation of a mineral statistics program.

The Economic Outlook for 1954. GERHARD COLM, *National Planning Associates.*

To maintain "full employment" next year an increase in total production of about $10 billion would be necessary. Surveys of present spending intentions of consumers, business, and government for the next year do not indicate any sharp increases or decreases in total buying but on balance a mild further decline. The defense demand of the Federal Government may be down by $2 or $3 billion next year, but the continuing rise in state and local expenditures for roads, schools, hospitals, and other improvements will offset much of the decline in Federal demand. Business enterprises are planning only slightly less spending for plant and equipment in 1954 than in the peak year 1953. According to a recent survey of consumer attitudes, people in general are optimistic about next year's income and feel that the present is a "good time to buy." However, some decline in consumer incomes is likely. Since there is no indication that consumers intend to spend a larger proportion of their incomes, consumer spending is likely to show a moderate decline. If one simply puts the fragmentary indications of present intentions into a coherent picture and allows for some decline in inventories, one would reach the conclusion that the year 1954 would bring a reduction in total production of about $10 to $15 billion instead of the desirable full employment increase of $10 million. This would mean a level of activity of 5 or 6 per cent below the full employment level. Unemployment might rise from 1½ million in 1953 to perhaps 3½ million in 1954.

Before accepting this as a "forecast" it must be recognized that present intentions, which are the basis for this outlook, may well be changed. In view of the possible weakness in markets for some goods, it is possible that business might engage in rather substantial inventory liquidation and might also revise downward its expansion programs. On the other hand, forward looking businessmen may think that conditions warrant a stepped up modernization program. They might also push ahead the development of new products and make the purchase of goods more attractive so that consumers are persuaded to use some of their liquid reserves for increased purchases. Finally, the Government in response to an economic downturn might adopt tax reductions beyond present plans, might adopt financial measures to stimulate the construction of residential houses, or might step up other useful programs. In our present state of knowledge it is only possible to indicate the economic trend on account of *present* intentions. This trend is mildly down. If the community maintains its confidence that the Government is ready and willing to act, it should be possible to prevent the downturn from developing into a depression. How consumers, business, and government respond to this trend, whether their responses will aggravate, mitigate, or reverse it, can be a subject of discussion but not of any forecast which pretends to be more than one man's opinion.

Economic Forecast of the Agricultural Situation, 1954. ORIS V. WELLS, *U. S. Department of Agriculture.*

No marked change in the domestic demand for food and other agricultural products appears likely in 1954 as compared with the current year. Also, foreign takings of United States farm products, while sharply reduced in the 1952–53 season from other recent years, appear to be at a level sustainable over the next year or so. Supplies of most farm products are expected to continue large in 1954 Carryover stocks may increase further by the end of the current marketing year, but a large part will be held by the Government. Acreage restrictions are likely to bring smaller wheat and cotton crops in 1954 and price support programs will continue to cushion the impact of large supplies on farm prices. With prospective conditions of demand and supply for farm products in 1954 approximately the same as in 1953, the average of prices received by farmers may hold near current levels. With cost rates to farmers stabilizing, the cost-price squeeze in agriculture is not likely to be intensified significantly in 1954.

Achieving Maximum Prediction per Unit of Testing Time. JOHN T. DAILEY, *Bureau of Naval Personnel.*

In an aptitude battery of finite length composed of pools of homogeneous items, how long and thus how reliable should each test be in order to maximize the composite validity of the battery and thus obtain maximum prediction per unit of testing-time? For tests composed of homogeneous items, test validity and reliability vary concomitantly with the number of items, and varying the test length alters

both reliability and validity in a predictable manner. Formulas are derived and presented for predicting correlations with tests of altered length.

When homogeneous items are added to a given test, the amount each successive item adds to the multiple validity falls off sharply. Items added last to a long test of highly valid type items may add less than items added first to a less valid type of test. Data are presented to predict the results of a shortening of each test in the Aviation Cadet Classification Test Battery. It is demonstrated that, for a given amount of testing time, appreciably greater multiple validity may be obtained by using a large number of short tests rather than a smaller number of long tests. These predictions were verified by an empirical study.

Family Interaction and the Transmission of Achievement-related Attitudes. FRED L. STRODTBECK, *University of Chicago.*

Forty-eight recorded discussions between father, mother, and adolescent son were analyzed and significant negative correlation was found between the power of the father in the family decision-making and the score of the son on an achievement-related attitude scale. The families were second generation (the son third), of Jewish and Italian ethnicity, divided between over- and under-achieving students and stratified into three socio-economic status groups forming a $2\times2\times3$ factorial design. The frame for selecting the sample was created by administering questionnaires to all children between 14 and 17 in parochial and public schools in an Eastern City. The study is a part of the research on the early identification of talented persons sponsored by the Markle Foundation through the Social Science Research Council.

The Validation of Testing Programs for University Students. WILLIAM B. SCHRADER, *Educational Testing Service.*

The widespread use of tests for selection and guidance in American universities has been accompanied by the growth of testing programs. A testing program may include all the steps needed in collecting data on test performance by examinees and in transmitting the test results in a convenient form to appropriate test users. Validation of tests for predicting academic achievement contributes to a program by aiding in the evaluation of program components, in the improvement of test offerings, and in the effective use of test results.

A comparative study of mathematical aptitude and achievement materials for predicting engineering school grades provided useful information for an administrative decision by the College Entrance Examination Board. A simple method of assigning optimal testing times to test parts indicated that the Law School Admission Test could be appreciably shortened while maintaining the same validity. Graphs have been used to aid in combining test scores with previous academic record and expectancy tables have been developed to aid test users in interpreting predictions. The planning and interpretation of validity studies should provide for: sampling institutions according to a defined plan, using test scores jointly with other predictors, studying homogeneous student groups, avoiding capitalization on chance, taking account of restriction of range of talent, using a regression approach wherever possible, and reporting results simply. Criterion development is a promising though difficult field. Broadening of the criterion to include other major outcomes of college and professional education than those reflected in grades is needed. Also needed is long-range validation of the tests administered in the 1920's against criteria of adult success.

Statistical Principles of Testing. JOHN MANDEL

The answer to many problems in science, both fundamental and applied, is found in small differences in the numerical values of a few measurements. In determining these differences, it is necessary to guard against biases introduced by the testing procedure. One of the principal functions of statistical design is to detect and neutralize these biases. Examples are given to illustrate this approach:

(1) In an experiment made for the purpose of testing the homogeneity of 4 batches of polyisobutylene by means of flow-vicosity measurements, a 4×4 latin square was used. Considerable day-to-day variation, as well as systematic effects of chronological order within days were found. Correction of the data for these systematic effects permitted an evaluation of the degree of homogeneity of the material within prescribed limits of uncertainty, that would not have been possible without statistical design. (2) In many cases experiments run in parallel display much greater agreement than experiments run at different times or in different laboratories. This fact can be used to increase precision by the well known device of the "control sample." Statistical methodology has broadened the idea of the control sample to include the concept of the statistical "block," thereby eliminating the need for an actual control sample in many instances. The idea is illustrated by a road test of eight automobile tire brands for rate of tread wear, run in accordance with a chain-block design with two-way elimination of heterogeneity. Both run-to-run variation and the effect of wheel position are eliminated from the comparison

of the brands. (3) In interlaboratory experiments, the specimens for test are usually allocated entirely at random among the various laboratories. A judicious utilization of the idea of the experimental "block" can sometimes increase the precision several fold. Sheets of vulcanized rubber often show considerably more variability than that observed on any one sheet. In an aging study of rubber, carried out as an interlaboratory experiment, the information per measurement might well have been increased six-fold if the specimens necessary to obtain individual aging curves had been taken from the same sheet

Some Comments on the Lot Plot Plan. L. E. MOSES, *Stanford University.*

The lot plot plan of sampling inspection was conceived by Shainin as a method to give a high degree of protection against acceptance of lots with fraction defective in the vicinity of 1/10 of 1%. The plan is supposed to be effective regardless of the character of the lot. It has found wide adoption in many industries in this country. Little is known of its theoretical basis. The plan calls for taking a random sample of 50 observations and plotting the histogram for the sample, as well as the mean and average range (from the 10 sets of 5 observations). From the appearance of the histogram the inspector decides whether to view the lot as "normal," "flat-topped," "long-tailed," "bimodal," "skew," or "truncated." For each such type there is a somewhat different way of deciding upon the acceptability of the lot.

This paper presents indications that the operating characteristics of the plan depend markedly on the character of the lot; that there is little hope of using a 50-observation histogram to detect "small" departures from normality which greatly distort the O C. curve. The special types of analysis prescribed for skew and bimodal samples are considered, and found to have rather unsatisfactory properties in general.

Continuous Sampling Plans. HARRY WEINGARTEN.

Continuous sampling plans (for acceptance inspection of material not assembled in "lots") available in the literature, were designed for a manufacturer interested in keeping a check on his production and also guaranteeing an AOQL. The plan proposed by H. F Dodge, "A Sampling Inspection Plan for Continuous Production" in the September 1943 issue of the *Annals of Mathematical Statistics*, is such a plan. Dodge's plan requires the alternation of 100% inspection and sampling. Two quantities i and f are specified: if after starting with 100% inspection, i consecutive items are found free of defectives, sampling at the rate f begins. If a defective is found during sampling, the inspection returns to 100% until again i consecutive items are free of defectives, etc. For a fixed value p of the incoming per cent defective, i and f determine the AOQL.

Emphasis in adopting the Dodge Plan for use by a purchaser (Bureau of Ordnance, Navy Department) was placed on reducing the amount of inspection required. If poor quality were submitted for inspection, the purchaser would, in effect, perform a screening operation for the manufacturer. The adaptation utilized the rejection number concept of lot acceptance plans; thus the Dodge Plan is operated as long as less than $a+1$ defectives are found. When inspection is interrupted, upon finding the $a+1st$ defection, it is not resumed until the manufacturer finds and removes the cause for defective product.

A sampling plan in Navord Standard 81 is defined by four quantities, N = anticipated volume of production, i = length of 100% inspection requirement, f = sampling rate, and a = maximum number of defectives allowed before interruption. This generalization of the Dodge Plan (Dodge Plan: $N = \infty$, $a = \infty$) produces mathematical problems of great complexity. If the effect of a plan defined by N, i, f, and a is described by

$$L_p = Pr \text{ (not interrupting inspection)},$$

it was possible to find L_p exactly, only for $a = 0, 1, 2$. For larger values of a an approximation was used.

Problems of Coordinating the United States Statistical System. STUART A. RICE.

The statistical system of the United States embraces many official, semi-official, and unofficial agencies and instruments, among which some degree of coherence is maintained by item to item adjustments among related tasks and processes. The relationships may be those of conceptual congruity or those of consistency in operational patterns and sequences. The process of maintaining coherence is called coordination. In the decentralized statistical system of the United States a particular problem which falls to the Office of Statistical Standards as the Government's central agency of statistical coordination is that of representing the general interests. These often override or lie between the interests and responsibilities of particular statistical agencies. Although a cross-cutting "statistical budget" is developed by OSS each year, its components must be fitted into the over-all budgets of departments and agencies, thus raising special problems. Other important problems of coordination are the establishment of balance among separate agency programs, of establishing lines of demarcation between governmental and non-governmental responsibilities and of finding the appropriate boundaries to the conceptions of "confidentiality" and national security interest in data. Finally, the problem of liaison between Federal

statistical agencies and the statistical profession is being solved in piecemeal and modest but satisfactory ways through advisory mechanisms established by the American Statistical Association.

Problems of Co-ordination in the Canadian Statistical System. HERBERT MARSHALL.

Canada's statistical system is highly centralised. The Dominion Bureau of Statistics was set up under a Statistics Act in 1918 and instructed to "organise a general scheme of co-ordinated social and economic statistics pertaining to the whole of Canada and to each of the provinces thereof." To reach this objective several types of co-ordination were necessary. Since Canada is a federal state the raw materials for some kinds of statistics are derived from the administrative records of both federal and provincial departments. These administrative records have to be fitted into the general overall scheme. Co-ordination with provincial departments is achieved mainly by annual or less frequent Dominion-Provincial Conferences in various statistical fields. The bulk of the Bureau's output is based on information collected directly rather than from administrative records and includes censuses of population, agriculture, industry, distribution, labour and prices, and numerous other fields. Co-ordination with those who fill in the questionnaires and those who use the data involves close liaison with numerous business and other organizations and a wide variety of users of the data. Co-ordination of the work done within the Bureau of the fourteen Divisions and numerous sections is of great importance. Uniformity in concepts, definitions, classifications, avoidance of duplication, and other aspects of co-ordination require constant vigilance on the part of interdivisional committees, and other supervisory efforts.

Methodological Problems and Findings of Study of Recipients of Old-Age Assistance. THOMAS G HUTTON, *Department of Health, Education, and Welfare.*

The major methodological problem faced by the Bureau of Public Assistance in making the study of old-age assistance recipients, as in all its studies, is how to collect data that are comparable from State to State when the Bureau must conduct its studies with personnel of 53 State public assistance agencies each operating under its own, and frequently diverse, policies, and with case workers in over 1,000 governmental units scattered throughout the length and breadth of the land. These workers have diverse backgrounds and varying degrees of technical skill. Contacts with State personnel and local case workers are almost entirely through the written word, and the Bureau is dependent on obtaining the cooperation of these workers in order to obtain data that are reasonably accurate. The solution to this problem while not a statistical one, but one in the field of human relations, is an integral part of making such a study. The paper discusses how this problem was solved.

The study of old-age assistance recipients was designed to obtain answers to two questions: (1) Who are these old people that are receiving public assistance, and (2) How do they live. Answers to the first question were obtained in terms of certain social characteristics; age, sex, race, marital status, and physical and mental condition. Answers to the second question were obtained through information on living arrangements, extent of home ownership, certain housing characteristics relating to extent of overcrowding, sanitary facilities in the home, and the use of modern conveniences such as electricity, the telephone and refrigeration. The degree to which children contributed to the support of aged parents, the amount and sources of income other than assistance, and the total amounts of income on which these people lived also contributed to answering the second question. The findings of the study on each of these points are presented in brief compass in the paper.

Methodological Problems and Findings of Survey of Aged Beneficiaries of Old-Age and Survivors Insurance. EDNA C. WENTWORTH, *Social Security Administration*

In the fall of 1951 the Bureau of Old-Age and Survivors Insurance conducted a Nation-wide survey of the economic resources of retired worker and aged-widow beneficiaries. The paper discusses the methods used in evaluating the economic situation of the beneficiaries and some of the findings for those who received benefits throughout the survey year. Slightly over three fifths of the beneficiaries had income in addition to their own independent retirement income, or they used assets. The chief sources of the additional income were public assistance, contributions from relatives outside the household, and earnings, usually from short-time employment. Two thirds of the beneficiaries either received public assistance or had less money income from all sources than public assistance would allow its recipients who lived alone in rented quarters; a sixth were assistance recipients. Half of those with the low incomes and no public assistance got along because they lived with relatives and were partially supported by them. Some of the others had noncash income of various kinds. Altogether, half the beneficiaries were partially dependent during the survey year, more of them on relatives than on public assistance.

If they used one tenth of their liquid assets each year, most (70 per cent) of the retired workers with the highest benefits—$60 to $68.50—would have independent retirement funds for the next 10 years of at least $900 a year if single and $1,500 if married Only 36 per cent with benefits of $50 to $59 would

have such independent retirement funds. Those with smaller benefits were worse off. Homes were owned by 45 per cent of the beneficiaries but the homes are not taken into consideration in this appraisal of beneficiary resources.

Intervals Between Onsets of Multiple Cases of Poliomyelitis in Families. ARTHUR S. LITTELL and GEORGE V. SMITH, *Western Reserve University.*

Recently P. E. Sartwell has shown that observed incubation periods of poliomyelitis are described by a log-normal frequency distribution. Using this distribution of incubation periods an expected distribution of intervals between onsets of pairs of cases exposed simultaneously was computed. Six observed series of intervals between onsets of initial cases and onsets of subsequent cases within families, 1310 intervals in all, are compared with the distribution expected if all cases in each family resulted from common exposure. The agreement between observed and expected distributions is very good in some of the series. In the others, the disagreements are opposite to what one would expect if there were an appreciable proportion of secondary infections within families. Instead of an excess of intervals longer than 7 days, which would represent secondary cases, there is a deficiency.

It is shown that, even if non-susceptible carriers are partially responsible for the spread of poliomyelitis, an observed distribution of intervals should show an excess due to secondary cases of intervals longer than 7 days over the number expected on the basis of common exposure. This argument points to one of two conclusions: (a) when multiple cases of poliomyelitis appear in a family, the source of infection is common, and spread of the disease from case to susceptible within the family is rare; or, (b) a ase of poliomyelitis can be infectious very early in its incubation period.

Unsolved Problems of Experimental Statistics. JOHN W. TUKEY, *Princeton University.*

It would not be misleading to say that there is only one unsolved problem of experimental statistics —"How can we identify the problems of experimental statistics?" (We can identify a good many unsolved problems by accident, but we probably miss many important ones for far too many years.) Experience to date indicates that difficulties in identifying problems have delayed statistics far more than difficulties in solving problems. This seems likely to be the case in the future, too.

Thus it is appropriate to be as systematic as we can about unsolved problems. Any system may be a start toward a partial solution of this one central unsolved problem. We shall try to do this by stating first some hypergeneral principles and then some general consequences. We shall strive to phrase these as generally as possible, in the hope of prolonging their useful life. The discussion of examples of these 18 general principles will set forth a number of unsolved problems, while a list of 37 provocative questions poses many more. The account closes with a discussion of the possibility of orienting experimental statistics toward problems rather than techniques.

Some general principles. If we feel that the detailed problems of experimental statistics arise from the interaction of certain general principles among themselves and with classes of experiments, it is reasonable to try to state and illustrate some of these principles. Most of these hang on four hypergeneral principles, which may seem harmless until we come to their consequences, namely: (A) Different ends require different means and different logical structures. (B) In each area, statistical method must and does evolve, mainly by adding *both* immediate ends *and* considerations. (C) While techniques are important in experimental statistics; when to use them and why to use them are more important. (D) In the long run, it does not pay a statistician to fool either himself or his clients.

Characterization of Distribution-free Statistics. Z. W. BIRNBAUM.

1. *Definitions.* Let Ω and Ω' be families of cumulative probability functions. A real quantity $W = S(X_1, X_2, \cdots, X_n, G)$ is a *statistic in Ω with regard to Ω'* if, for any $G \epsilon \Omega$ and $X_1, X_2, \cdots, X_n$ in the n-dimensional sample space E_n of a random variable X with the c.p.f. $F \epsilon \Omega'$, this quantity 1° is defined almost everywhere in E_n and 2° has a probability distribution; this probability distribution is then denoted by $\mathcal{P}[S(X_1, \cdots, X_n, G); F]$.

If $\Omega = \Omega'$ and the statistic $S(X_1, X_2, \cdots, X_n, G)$ has the property that $\mathcal{P}[S(X_1, \cdots, X_n, G); G]$ is independent of G for $G \epsilon \Omega$, then $S(X_1, \cdots, X_n, G)$ is a *distribution-free statistic in Ω.*

Let Ω be a family of c.p.f.'s such that the inverse function $G^{(-1)}$ can be defined for each $G \epsilon \Omega$. A statistic $S(X_1, \cdots, X_n, G)$ in Ω with regard to a family Ω' is called *strongly distribution-free* if $\mathcal{P}[S, X_1, \cdots, X_n, G); F]$ depends only on $\tau = FG^{(-1)}$ for all $G \epsilon \Omega$, $F \epsilon \Omega'$.

If there exists a function Φ defined on the n-dimensional unit cube and symmetric in its arguments such that for any $G \epsilon \Omega$, $F \epsilon \Omega'$ we have $S(X_1, \cdots, X_n, G) = \Phi[G(X_1), \cdots, G(X_n)]$ almost everywhere in E_n for the random variable X which has c.p.f. F, then $S(X_1, \cdots, X_n, G)$ is a *statistic of structure* (d) in Ω with regard to Ω'.

2. *Problems.* Some types of problems which may be formulated in terms of the concepts defined above are: (a) given Ω, characterize the class of statistics which are distribution-free in Ω, or strongly dis-

tribution-free in Ω with regard to a given Ω′; (b) given Ω and a loss-function L, determine the distribution-free statistic which is optimum for a specified decision-theoretical problem.

The following are some results obtained by the speaker and H. Rubin for problems of type (a): If a statistic in the class Ω_1 of all continuous c.p.f.'s with regard to Ω_1 has structure (d) then it is distribution-free in Ω_1; but not every statistic of structure (d) in Ω_1 with regard to Ω_1 is distribution-free in Ω_1. A statistic in the class Ω^* of strictly increasing continuous c.p.f.'s with regard to Ω^* is strongly distribution-free if and only if it is of structure (d). For problems of type (b) O. P. Aggarwall has recently shown that for estimating a cumulative probability function the minimax procedure invariant under all permutations of the sample is, for certain plausible loss functions, either that based on Kolmogorov's statistic or on some modifications of that statistic.

Inadequacies of the Construction Estimates as General Economic Measures. Walter E. Hoadley, Jr., *Armstrong Cork Company.*

Construction merits general recognition as a major American industry, but few leaders within this diversified industry show much interest in aggregate activity statistics, and most "outsiders" have so little first-hand knowledge of construction that they tend to accept almost without question any available measures of construction, especially from "official" sources.

At least five general inadequacies stand out in the construction statistics field at present: (1) a widespread lack of understanding and appreciation of the importance of adequate construction statistics to public and private policies, (2) weaknesses in the basic construction series currently available, (3) major gaps in construction statistics, (4) the absence of a comprehensive plan to strengthen government statistics in the construction field, and (5) insufficient funds to insure suggested improvements can be carried out successfully. As construction activity has grown in size and importance to the whole economy, our knowledge about it has lessened. This unfortunate condition is not due to lack of competence in the governmental statistical agencies, which are doing just about the best job that can be done with the resources available. It is due primarily to the failure of Congressmen, businessmen, and others to recognize the importance of sound data as a basis for market planning or to recognize just how inadequate the present data are for this purpose.

Since the Korean War, for the first time a construction estimate series—the nonfarm housing start series of the U. S. Bureau of Labor Statistics—was written into law as a "trigger" statistic to help time changes in a government control (i.e., residential credit). This last move, perhaps more than any other single development, stimulated new interest in government statistics across the home building industry. Government construction specialists also quickly recognized the need to reexamine the adequacy of their statistics in light of their increased role in policy determination. Nevertheless, few major policy decisions affecting the construction industry can now be properly made on the basis of the statistics available.

The most encouraging development this year has been the extent to which policy level people within and outside government have become aware of many deficiencies in construction statistics and have started to promote plans designed to improve several key series.

The three principal statistical gaps are: (1) Current indications of changes in the national housing inventory, reflecting the continuing impact of new buildings, conversions, and demolitions. (2) Estimated vacancies in old and new dwellings. (3) Estimated expenditures for "fix-up" purposes (i.e., repairs, additions, and alterations to existing construction). These data are absolutely essential for the housing and construction segment of the economy to prosper and avoid major swings.

The pressing job to be done in government—with the aid of all persons interested in construction—must be to obtain some general agreement as to the basic objectives of government in the construction field. Government policy officials must decide what *they* feel *they* need to make sound decisions in the interest of growth and stability in the ecomomy. Only when the major deficiencies and gaps in "public policy" statistics on construction have been substantially eliminated should attention be directed to meeting the numerous and ever-present requests for more detailed market type information—except of course as such becomes available as a by-product of public policy statistics. In my view, the construction industry and the economy as a whole have much more to lose from inappropriate public policies than from the absence of fairly detailed information about particular segments of construction markets.

There is reason to believe that construction statistics programs will be strengthened in the months ahead—not simply because of our wishes, but rather because "public policy" construction statistics—on their own merits—promise to be recognized more widely as far too important to public policy determination to be allowed to remain inadequate or to deteriorate further.

Analysis of Variance Models in Sedimentary Petrology. J. C. Griffiths.

Investigation of the petrography of sedimentary rocks by analysis of variance already embraces completely randomized, randomized blocks, unbalanced lattice and factorial designs. The hierarchy of petrographic units represented by rock-types, formations within rock-type, outcrops within formations,

specimens within outcrops and aliquots within specimens leads to multistage sampling patterns and completely randomized designs.

Interacting factors and randomized blocks design may be introduced by using two or more spatial dimensions analogous to "rows and columns" of agronomic experiments. In contrast, however, the aim of such arrangements in the analysis of sediments is the evaluation, and not solely the segregation, of this "fertility gradient". As another interacting factor order may be introduced either as a stratigraphic variable or as a sequential arrangement of a series of experiments. Equivalent to the "variety" effect in agronomic investigations, observers performing the same experiments, leads to evaluation of operator variation. In this connection operator inconsistency or "discrepance" has proved of vital interest in setting the sensitivity level of many petrographic investigations.

Heterogeneous variance has complicated many of the comparisons but by estimating components of variation it has been shown that mean differences are, perhaps, less important than differences in variability in characterizing sedimentary rocks and the processes by which they are formed. Equal sampling based on the assumption of equivalence of classes, as for example in using, arkose, graywacke and quartzite, has proved fallacious and introduces heterogeneous variance. Sampling in proportion to variability appears to be the obvious solution as emphasized by some recent analyses of variance of sphericity and roundness of quartz grains in sediments.

On Testing the Association of Mineral Occurrence with a Set of Observable Characteristics. RAY MICKEY.

Let a topographic map M be given for an area to be explored for the presence of a set of minerals. Let V be the vector function defined over M which associates with each point of M a set of values of geological characteristics, and let Z be the function over M which assumes the value one for those points whose mineral density is sufficiently large and which assumes the value zero otherwise. Let x be a random point in M whose distribution has constant density over M. We will say that V and Z are associated if $V(x)$ and $Z(x)$ are not independent chance variables.

We consider the possibility of testing H_0, the hypothesis of no association, by the use of conditional tests of the following class. Let $t = t(V_1, \cdots, V_n, Z_1, \cdots, Z_n)$ be a test statistic with observed value t_0. Conduct a series of sampling experiments in which $V_1, \cdots, V_n$ and $\sum Z_i$ are held fixed and H_0 is assumed. The test is conducted by the use of a binomial sampling plan. Perhaps the simplest of these of level $1/k$ are: reject H_0 if

$$\sum_{i=1}^{mk-1} u_i \leq m - 1,$$

where $u_i = 0$ if the value of t obtained from the ith sampling experiment is less than t_0 and $u_i = 1$ otherwise

An example is presented in which use is made of the statistic

$$t = \max_{i \in S_0} \min_{j \in S_1} \rho(V_i, V_j)$$

where S_α is the set of indices for which $Z_i = \alpha$, and ρ is a suitable metric.

Problems in Sampling the Phosphoria Formation. R. A. GULBRANDSEN and V. E. MCKELVEY, *U. S. Geological Survey*

The Phosphoria formation of Permian age contains vast resources of phosphate and other materials distributed over an area of 135,000 square miles in eastern Idaho and adjacent states. In an effort to appraise the total mineral resources in the formation and to determine its origin, the U. S. Geological Survey has smpled the full thickness of the phosphatic members of the formation at about 200 scattered localities. The sampling plan was based mainly on geologic intuition, or geologically educated judgment; but it took account of previous information available on the variation in composition and thickness of the beds, the distribution of the rocks and their outcrops, and the purposes the samples were expected to fulfill.

All samples collected are being analyzed for phosphate and other constituents of prime economic importance. It is expected that the average of five to fifteen widely spaced samples collected from an area of a few tens of square miles will represent the grade and thickness of deposits in that area and that the range in grade and thickness in any one area will indicate the magnitude of variation to be expected there. The chief uncertainty in estimates of reserves will result from lack of information on the presence or absence of beds at depth. Some of the samples will be analyzed for constituents or properties of scientific interest only. These samples will be selected randomly from samples grouped according to rock-type, stratigraphic position, or area.

Longitudinal Studies of HIP Experience, Background and Some Findings of Pilot Study. Neva R. Deardorff. *Health Insurance Plan of Greater New York.*

The longitudinal studies now in progress at the Health Insurance Plan of Greater New York are essentially examinations of the experience of people covered for comprehensive prepaid medical care for periods of time up to a maximum of four years. Three phases of the experience of these people are being examined: their enrollment experience, their utilization, and the diagnoses reported for them. The happenings to a constant group of people can be reported from year to year and the changes noted in the *group's* pattern of experience, irrespective of the sequence of changes experienced by any one person or family. Another focus of interest is the *movement* of persons and families from one category to another over the years.

Longitudinal studies point toward the treatment of each person's *history* rather than his condition at a given time, as the unit of statistical manipulation. In a sense, it is the statistical approach to a group of persons or families, each with a continuing life history to be considered, rather than the construction of cross-section views of these people's characteristics made from time to time. This Project was planned and is being carried out under the direction of a committee composed of persons who have no affiliations with HIP, implemented by a specially employed staff. The work of the Project is financed by grants from the Commonwealth Fund and the Rockefeller Foundation.

Among the requirements for effective longitudinal studies is the need for a comparatively large number of subjects since almost all groups are subject to attrition from death and other causes of loss of subjects and since patterns of change tend to proliferate with each added year of observation. (A diagram illustrating these points was presented.) The HIP studies are based on random samples comprising 8,025 persons who were under observation from 1948 to the end of the study period (December 31, 1951), 9,325 since 1949, and 3,500 since 1950. Besides these, there are 8,000 persons whose coverage was terminated or had been interrupted or who were spouses and children acquired in the last year (1951) of the study period. These 8,000 cases have been included not only because of their place in annual tables, but also for purposes of some special analyses on enrollment, utilization, and diagnostic conditions. Altogether a total of 28,850 persons are included in these studies. Three tables from a pilot study were distributed to illustrate a few types of analysis which are being employed in the presentation of data.

Longitudinal Study of Health Insurance Plan of Greater New York. Nathan Goldfarb, *Health Insurance Plan of Greater New York.*

Although there is an increasing recognition of the particular advantages of longitudinal data as contrasted to cross-sectional data, there has been only limited experience thus far in the organization and processing of large masses of data for longitudinal analysis. It would, therefore, be extremely useful for those who are interested in longitudinal analysis to discuss the problems and methods of organizing and analyzing longitudinal data.

There are two basic sources of data for this Special Research Project, namely, (1) the documents completed by the physician and submitted to HIP at regular intervals, which identify the individuals and give the tentative diagnosis and other related information, and (2) the registrar records which provide the enrollment data and the personal identifying characteristics for each individual. In order to determine whether the physicians' reports were being prepared with sufficient care for the needs of the Research Project, the data from these reports were checked for correspondence with the records in the physicians' files and found satisfactory. The 10 per cent sample of families and individuals covered by the HIP for which all the Project data are being obtained, was also tested and found representative for a number of characteristics.

The objectives of the Project were framed in a series of questions. It was evident that a history and summary of the medical experience for each individual for the entire period of the study was required. The Project then set about to summarize data covering approximately 30,000 people, with about 315,000 medical services, for a period ranging up to four years. Coding and tabulating methods were evolved which permitted an examination of the medical experience and care for each person in the sample. Codes were developed to give the detailed diagnoses within each study year, cumulative diagnostic codes to summarize the total diagnostic experience for an entire year, and still another set of cumulative codes to summarize the diagnostic experience for combinations of successive years of coverage—2 years, 3 years, and 4 years. Chronic diseases were classified separately from the non-chronic diseases. Every medical condition was coded, but there was no segregation of "primary" and "secondary" diagnoses, nor was there any count of episodes. These decisions were in part dictated by the nature of the reporting. Where a diagnostic description was evolved over a period of time, the "best" diagnosis was chosen on the basis of a set of rules which, for example, gave priority to the specialist's report over that of the general physician, and to the later diagnosis over an earlier one in a sequence of visits to the doctor

The tabulations on utilization will show the trend for various cohorts of individuals indicating how, with the passage of time, the factors of age, sex or a diagnostic type affect the quantity of medical care.

"Spot Checks" in Lieu of Complete Censuses. HOWARD C. GRIEVES, *Bureau of Census.*

In the summer of 1953 the Congress indefinitely postponed the Quinquennial Censuses of Manufactures, Mineral Industries, Business and Transportation. It substituted the sum of $1.5 million for "spot checks" in the fields of manufactures, business and agriculture. The appropriation is being used for three broad purposes: (1) to fill some of the gaps in national aggregates created by the deferral of the complete censuses; (2) to increase the timeliness of reports; and (3) to develop some new statistical indicators.

Specifically, sample surveys will provide national totals of wholesale, retail, and service trades for the year 1953. The Annual Survey of Manufactures for 1953 will also be carried out. Monthly estimates of retail sales and the labor force are being improved by increasing the number of primary sampling areas. The monthly wholesale series will henceforth utilize a probability sample developed for the 1953 Annual Survey. The publication *County Business Patterns* is being compiled on a greatly accelerated time schedule covering 1953. Advance estimates of retail sales by major kinds of business employing a probability sample are being released on the 10th day following the month covered. A complete census of retail trade, using an abbreviated report form, is being taken in Dallas on a trial basis with financial assistance from local business. Finally, an effort is being made to improve the monthly measures of retail inventories. A number of other smaller projects are also in progress.

While much useful data will be forthcoming it is important to note some of the purposes of complete censuses which cannot be achieved with the resources available. The most important omission will be statistics for small areas, cities, counties, etc. Similarly, detailed characteristics, fine industry breakdowns, commodity statistics, etc., cannot be provided. The planned improvement of the samples in current use based on enumerations of the entire universe correctly classified must be deferred. Similarly, the planned coordination of Federal statistical series to achieve uniform classification is postponed until the complete Censuses are taken.

Developments in Collection and Processing of Mass Statistical Data. MORRIS H. HANSEN and WILLIAM N. HURWITZ.

Large-scale census and survey operations offer numerous opportunities for the application of statistical and other scientific methods. Such applications are now frequently referred to as operations research. The general aim in developing and improving methods is to produce the needed results at minimum cost, and within specified time schedules and other restrictions. It involves making the fullest possible use of available resources.

Modern sampling methods make it possible to provide timely results of known precision at low cost through monthly, quarterly, annual, or special sample surveys of the population, retail trade, manufacturing, and other subjects. Also, sampling methods are applied in the major censuses to collect and tabulate much of the basic information at reduced cost, to produce more timely results, to check on the quality of the census results, and in other ways.

Important uses of statistical methods in connection with both censuses and current work is in the control of quality. Quality control usually consists of a system of sample inspection and process control introduced to insure at low cost that the final product will meet certain quality standards. Quality check operations are to be distinguished from quality control. Quality checks have been introduced in situations where it has not been feasible to control quality. They involve the use of intensive measurement methods on small samples, and have been used especially in evaluating the quality of the field work in major census operations. Not only do they guide in the proper use of the results, but they provide the basis for improving the design of future censuses and surveys.

Other areas for the application of statistical methods include evaluation of alternative procedures through properly designed experiments, such as alternative types of questionnaires, collection processes, or training methods. Also, the effects of editing and other operations have been investigated and modified or mechanical techniques substituted. Our work has indicated that much editing formerly regarded as necessary, can be eliminated or reduced with only a trivial effect on the accuracy of results.

Another area for application of statistical methods is in setting standards or incentives for personnel performance. Very substantial production gains have been achieved simply by administrative incentives through performance standards set on the basis of appropriate samples of work and sample observations of workers.

Large scale electronic computing equipment is being applied and will have a rapidly increasing role in the future, as will, also mark sensing or reading methods. The introduction of such equipment, together with sampling, quality control, techniques for measuring and controlling response errors, and other methods mentioned should result in substantially increasing the timeliness of major census results

as well as reduce their costs. The application of statistical methods in the census has been the principal work of a small group of versatile scientists interested in applications. They are trained in mathematical statistics, survey methods, psychology, and other techniques, and are given complete authority to ask questions and investigate and recommend, but are kept free of operating and responsibilities.

The Current Statistics Program of the Census Bureau. A. ROSS ECKLER and CONRAD TAEUBER, *Bureau of the Census.*

Much of the current statistics program of the Bureau of the Census exists to serve needs not met by the censuses, nor by reports which are a byproduct of government administration nor by reports based on special-purpose collections by government or private agencies. Some surveys and sample censuses develop national or regional estimates for items covered in the censuses; some surveys provide essentially complete coverage for a segment of the universe covered by a census; and some current reports, like those on foreign trade, are based upon documents connected with government administrative programs. Sample surveys are of major importance in meeting the needs for current data, and in relieving the pressure for inclusion of items in major censuses.

With allowance for increases in basic costs, the resources for current statistics programs in the Census have been reduced by more than 50 percent since 1947. Although program losses have been numerous, efforts have been made to retain key items and to apply improved methods, such as the introduction of a sample of 230 primary sampling units for the current field surveys in place of the 68 area sample used previously. The Bureau's resources for collecting and tabulating statistical data are used not only for carrying out the Bureau's own program, but also in performing similar tasks for other governmental units and for private agencies, when proper safeguards concerning the public use of the results can be assured. Through the increasing use of data available from administrative records and through extension of services, where appropriate, to other governmental, as well as to non-governmental, agencies the Bureau of the Census may be successful in offsetting some of the losses in statistical information resulting from program reductions in recent years.

Preliminary Tests and Pool Rules. T. A. BANCROFT, *Iowa State College.*

Certain applications in experimental design and regression analyses involving preliminary tests of significance can be arranged as special cases of a test of a general linear hypothesis in canonical form incorporating a preliminary test of significance. In order that one may have confidence in the use of such analyses it is necessary to investigate possible biases in estimates and the power and size of such complex tests. Bechhofer, using exact methods, was able to evaluate the integrals involved in the expression for the power and size of such tests in closed form for only a few special cases of the parameters involved. In order that information on power and size of such tests be explicitly available, certain approximate formulas have been developed by the author by making use of a central chi-square approximation to the non-central chi-square distribution. Empirical studies indicate that the approximate formulas give usable results, in particular for important values of parameters which have been so far unattainable by exact methods. Investigation of exact error bounds are in progress.

An Extension of Preliminary Tests for Pooling Data. D. V. HUNTSBERGER, *Iowa State College.*

Mosteller (*Jour. American Stat. Assn.*, Vol. 43, 1948) has shown that, if $\bar{X}_1$ and $\bar{X}_2$ are independent estimates for the means μ_1 and μ_2 of two normal populations with the same known variance, the "sometimes pool" estimator for μ_1 based on a preliminary test of the hypothesis $\mu_1=\mu_2$ is subject to large losses of efficiency, relative to $\bar{X}_1$, for some range of $\gamma=\mu_1-\mu_2$. In this paper it is shown that, if $T=(\bar{X}_1-\bar{X}_2)\sqrt{n}/\sqrt{2}\sigma$ is used for the preliminary test, the estimator $W(T)=\phi(T)X_1+(1-\phi)(\bar{X}_1+\bar{X}_2)/2$, where $\phi(T)$ is a continuous function of T, provides a marked reduction in the maximum possible loss of efficiency and extends the range $|\gamma|\leq C$ for which a gain may be realized. If $\phi(T)=1$ for $|T|\geqq T_\alpha$ and $\phi(T)=0$ for $|T|<T_\alpha$, where Prob. $(|T|\geqq T_\alpha)=\alpha$, $W(T)$ reduces to the estimator based on a preliminary test of size α. The efficiencies of $W(T)$ and the "sometimes pool" procedure are compared for a particular function $\phi(T)$. The results obtained for this example suggest that whenever a preliminary test for a possible pooling is indicated an estimator analogous to $W(T)$ will provide a substantial reduction in the maximum possible loss of power or efficiency of subsequent inferences.

Simultaneous Test of Linear Hypotheses by Analysis of Variance Methods. M. N. GHOSH, *University of North Carolina.*

In the analysis of variance for survey type of data, the tests for different hypotheses corresponding to sets of parameters are not usually independent. A notion of quasi-independence has been developed in this paper which depends upon the irrelevance of extraneous parameters in the first and second kinds of error for the test of each of the hypotheses. In the simultaneous test of the hypotheses, $H_1, \cdots H_s$, decisions of the type accept H_1 or reject H_1, accept H_2 or reject H_2, etc., are considered and

the significance level of such a test is defined as the probability of rejecting at least one of the hypotheses $H_1, \cdots H_s$, when all of them are, as a matter of fact, true. A method has been given for simultaneous test at a significance level α, when the least squares estimates of the sets of parameters for different hypotheses are non-orthogonal. These methods are applied to analyze growth data for children especially in relation to stage of sexual maturity and certain blood chemicals.

Multivariate Analysis of Covariance for a Latin Square. WILLIAM J. MOONAN, *University of Minnesota.*

It is known that, for some problems which are appropriately evaluated by analysis of variance and covariance methods, the procedures of univariate analysis must be generalized to the multivariate case. The reasons why such generalizations are necessary have been made quite explicit by H. Hotelling (1947) and others.

The multivariate analysis of variance and covariance of a 4×4 Latin Square is derived in this paper. The method of derivation follows the usual least-squares approach. However, the technique is followed for a matrix variate instead of a single variate. In particular, the multivariate normal model is assumed, the estimation space is specified, the normal equations defined, and the estimates of the parameters are made. Hypotheses on the multivariate parameters are given and the least-squares process is repeated to determine the sums of squares and cross-products which are appropriate for testing the hypotheses. A numerical example is carried through to illustrate all of the calculations necessary to evaluate the parameters and make tests of significance. The method of analysis is quite similar to the usual univariate calculations.

Weighting Coefficients for Age-Adjusted Death Rates. HARRY SMITH, JR. *University of North Carolina.*

The problem of finding a descriptive mortality index for a community is reviewed. The methods for adjusting death rates for age are discussed in terms of the criteria for the use of each. The advantages and disadvantages of each are also listed. A new method of adjusting death rates for age is proposed when the comparison of two communities is to be made. The assumption upon which this new method is based is that the mortality functions of the two communities are of the same shape and separated by a constant differential. Any overlapping that occurs is assumed due to sampling fluctuations. The solution is based upon the method of maximum likelihood. A comparison of the results using this new method with results of other methods is shown. The use of a discriminant function in determining those age groups most influential in discriminating between states in the past ten years is also studied as another possible means of calculating an over-all mortality rate.

Some Statistical Evidence on Economies of Scale. FREDERICK T. MOORE, *Bureau of Mines.*

Statistical evidence on economies of scale in manufacturing is scarce because of the necessity of compiling detailed cost studies on plants, or, on an engineering basis, computing production functions. In either case the procedure is tedious. Engineers have derived certain rules of thumb for evaluating the relationship between capital costs and capacity. One such rule is the ".6 factor" rule which states that the increase in capital cost is given by the increase in capacity raised to the .6 power. For example, the surface area of a spherical tank (which essentially measures the cost) increases as the volume of the tank (capacity) to the two-thirds power. The ".6 factor" has been applied both to process equipment and to complete plants. It is apt to fit best industries which are (a) capital intensive; (b) continuous flow operation; and (c) with a homogeneous standardized product. The chemical and mineral processing industries in general meet these criteria.

Tests were made on samples of plants from several industries, using the formula: $\log E = \log a + b \log C$ where E = capital cost; C = annual capacity and a and b are constants. Values of $b < 1$ indicate economies in capital cost and since operating costs tend to behave in the same fashion, this also indicates economies in scale. In the tests the following values for b were obtained: alumina .95; aluminum reduction .93; aluminum rolling .88; aluminum extrusions 1.0; cement .77; tonnage oxygen .63. Tests were also made for individual processes in these industries. A separate study of the production function in petroleum pipelines indicated economies of scale up to approximately 200,000 barrels of throughput per day and constant returns for larger throughputs.

It was suggested that a simplified method for evaluating economies of scale can be visualized in three steps: (1) extensive engineering information is available showing the cost of process equipment relative to some engineering magnitude (e.g., square feet of heating surface); (2) the engineering magnitude can be related to capacity by an appropriate formula (e.g., the capacity of a tank can be related to its diameter); therefore (3) by combining steps (1) and (2) above, cost can be related to capacity. This method can be applied to process equipment and the processes combined to show a complete plant. Since the method may be used for a sample as small as two, it appears to offer a simplified approach to evaluating economies of scale

Stability of Technical Coefficients: Evidence from Inter-Plant Differences in Labor and Materials Productivity. A. PHILLIPS.

Technical coefficients for Leontief matrices appear to be quite stable through time. This may arise from two distinct conditions. First, there may be a relatively stable industry technology which causes each plant to have approximately the same coefficient. Random influences affect the plant coefficients but tend to offset each other, resulting in stable aggregate industry coefficients through time. On the other hand, industry coefficients may be stable even with very different plant coefficients so long as the plant coefficients are themselves quite stable and the plants produce consistent shares of the total industry output.

Evidence from one industry shows the latter to be the case. Little central tendency of plant coefficients is found. But each plant tends to have the same inputs of labor and materials per unit output each year and produce the same relative share of industry output. Differences among plant coefficients appear related to other factors such as principal product type, integration and kind of material consumed, though most differences cannot be explained with available information. Since each plant may have its own technology which is properly part of the industry, technology is not ruled out as the cause of stability, though one fairly unique and stable industry technology does not appear to be the cause.

Adequacy of International Trade Statistics for Economic and Business Analysis. J. EDWARD ELY.

There are important conflicts between the various users of international trade statistics as to what the statistics should and can do. These conflicts have been resolved in different ways in different countries with the result that countries follow widely varying definitions and practices in compiling their foreign trade statistics. It follows from this that what may appear from the comparison of the statistics of trading partners to be errors in the statistics, are frequently differences caused by differences in definitions.

International trade statistics face a unique problem in the fact that, in contrast to all other types of statistics, more than one sovereign country compiles detailed information on the same phenomenon, in this case the movement of goods from one country to another. The detailed figures presented by a country in regard to its foreign trade are, therefore, subject to searching comparison with similar detailed figures released by another country on the same transactions. Adequate trade statistics for such multiple-country use will face major conflicts of interest with domestic uses. Since many of these and other conflicts may turn out to be irreconcilable it may be that there will have to be a substantial increase in the use of dual compilation procedures designed to fulfill conflicting needs for information.

Accuracy of Foreign Trade Statistics. ÖSCAR MORGENSTERN.

Accuracy of Foreign Trade Statistics can be estimated by comparing records for the same transaction by exporting and importing countries. This reveals for all periods very large differences in pairwise trade both for values and quantities. Difficulties in classification of commodities, transportation costs, tariffs can explain only part of these differences. Even if successfully enumerated, the various factors work jointly upon the final number record and their contributions are not known individually at each instant. Hence substantial corrections are not feasible. An exhaustive enumeration of factors does not make foreign trade data, when they have the properties listed by these factors, suitable for a fine-structured theory of international trade using such concepts as terms of trade or for monthly comparisons with exchange rates, interest rates, etc.

The most striking example is offered by gold, a clearly defined, easily recognizable commodity of high value relative to quantity, costs of shipment and usually accorded exceptional care in handling. Yet statistics of gold movement, pairwise compared, reveal disastrous differences whether 1900, 1907 or 1928, 1935 are taken as samples. This holds true for monthly as well as cumulative yearly data.

There appears to be little hope ever to reconstruct series for the past on gold useful for most purposes of economic theory. Present information can only be correlated with the discussion of broad tendencies in economic life. The same is true, *a fortiori*, for the far more involved statistics of other commodities and the large categories into which they are put. These sober facts should lead to the establishment of a program for future data on foreign trade with known errors, perhaps on a sampling base.

On Respondent-Nonrespondent Differences Observed in the Pittsburgh Morbidity Surveys. DANIEL G. HORVITZ, *North Carolina State College.*

In morbidity surveys in particular, and other surveys as well, it is not uncommon to find a field procedure in use which permits the interviewer to obtain information required on *all* members of each selected household from any responsible member who happens to be available at the time of the call. The efficiency of this procedure is open to question whenever the information supplied by the respondents on the other individuals (nonrespondents) is inaccurate.

In the absence of an opportunity to measure the response bias directly, data collected on illnesses experienced in the month prior to interview in two surveys (carried out a year apart with a probability sample of approximately 3000 households located in the Arsenal Health District of Pittsburgh) were subjected to internal study in an attempt to characterize the rather large differences in the rates observed for respondents and nonrespondents. The data were examined first to determine the relationship, if any, of these differences to age, sex and size of household. Almost in general the data yielded no indication of a relationship between the observed respondent-nonrespondent differences and these variables whether for all illness or specific categories of illness. There was slight evidence of a greater respondent-nonrespondent difference (i) for males than females for noninfectious conditions and (ii) for those in the 15–34 year age group versus older age groups for all types of illness. This latter observation was accounted for by a strong relationship between respondent-nonrespondent differences and age for female conditions associated with pregnancy and childbirth. For these conditions the respondent-nonrespondent differences can be considered actual rather than due to any reporting bias. The respondent-nonrespondent prevalence rate differences observed for the specific chronic conditions of arthritis and heart disease also were found to be unrelated to age, sex and size of household.

Examination of the respondent-nonrespondent differences for illnesses still in progress on the date of interview versus those which had already ended indicated that a reasonable portion of the total difference may be an actual fact. This observation together with the consistency of the differences with age, sex and size of household requires further verification from morbidity studies conducted elsewhere. They also require (pending substantiation) investigation of the feasibility of a scheme for morbidity surveys whereby in a relatively small random segment of the sample households the respondent is selected at random from those eligible. The respondent-nonrespondent differences in these households should provide an estimate of the reporting bias portion of the differences observed in the remaining households. This suggestion presumes that the actual and bias portions of the differences do not vary in a compensating fashion to yield total differences unrelated to age, sex and size of household

Questionnaire Design and Related Methodological Problems in the Canadian Sickness Survey. ROBERT KOHN, *Dominion Bureau of Statistics, Ottawa.*

Problems of design are somewhat different for questionnaires and schedules, although many requirements apply to both. The basic document in the Canadian Sickness Survey was an interview schedule. Principles of design are discussed in three main categories: 1. contents, 2. wording, and 3. layout and format. The solution for particular problems will always depend on the type of informant and enumerator to whom the questions are directed.

The contents are dictated by the objective of the Survey and the tabulation program. They should be kept to the minimum required. If after the beginning of the survey the need arises to add supplementary questions, these should not be allowed to jeopardise the original ones and enumerators should be fully informed regarding their purpose and meaning. The wording must be such as to elicit the same answer in the same situation from a variety of people. It must be adapted to those of the prospective informants and enumerators who know and understand least about the survey. It must be easily understandable and yet concise. Instructions must also be drafted with great care and some machinery must be provided for prompt and uniform explanation and supplementation of the instructions if required during the survey. Layout and format must be such as to facilitate interviewing, recording, checking, editing, and coding. Pretesting of schedules will reveal shortcomings before the actual survey gets under way. Qualification and training of enumerators are important and so are methods to solicit and maintain the co-operation of the informants.

The Future of Railroad Shares. PIERRE R. BRETEY, *Heyden Stone & Co.*

During the past decade the rails have suffered from three major factors: (1) Delays by regulatory authorities in adjusting freight and passenger rates to higher costs, particularly following sizable wage increases. There have been many such delays since World War II. However, the "adjustment lag period" has consistently been reduced over recent years, and may be almost entirely eliminated through passage of legislation. (2) The inflexibility of a large percentage of railroad costs, with the result that historically earnings have declined sharply when gross revenues contracted. Notwithstanding this inflexibility, a number of railroads had achieved a flexible cost position before the roll-over adjustment of 1949 and demonstrated better expense controls than exhibited by many industries. (3) A lowered institutional appraisal resulting from the collapse in railroad earnings and in railroad equities both in 1929–1932, and again in 1938, collapses not duplicated since those two periods. Today most investment railroad equities are selling at from only four to five times estimated 1953 earnings, and railroad managements agree that individual properties are in the best physical condition in their history and that operating efficiency of most railroads has been restored following the rapid increase in costs immediately following

the end of World War II. They are now entitled to be regarded as akin to many other industries sensitively dependent upon the strengthened ability of individual managements in adjusting to changing conditions. I confidently expect that budgetary controls, maintenance cutbacks and more efficient operations resulting from cumulative benefits of some $9 billion of capital expenditures made since V-J Day, will maintain 1954 earnings of the Class I railroads at between $600 and $700 million. We are confident that after the lapse of fifteen years or more, institutional investors will once again reenter the railroad market. If this demand does make its appearance, it may well serve to spark another major market leg in the long road toward ultimate restoration of railroad credit to its once former high estate

The Changing Stock Market. Sidney Lurie, *Paine, Webber, Jackson & Curtis.*

There are a number of inter-related factors which have an important bearing on the outlook for our economy and, in turn, the stock market: (1) The population growth already experienced and still in prospect is opening up new markets for American industry—for the gains are in the ranks of the consumer rather than producer. (2) The redistribution of our national income, with the greatest increase being in that segment which proportionately spends the most, injects a new stabilizing force. (3) America's inventive genius in developing new products which open up new markets has spelled profits for the investor as well as manufacturer. While these considerations give rise to an era which historians may term "The Fabulous Fifties," two other factors have created an enigma for the security buyer: (A) Although common stocks today are more respectable than ever before in Wall Street history, the public has lost sight of the fact that the rewards always go to the risk-taker. The security buying public has forgotten there is nothing illegal, immoral or fattening to speculation—that there is little basic difference between intelligent speculation and intelligent investment. (B) The big financial news of the postwar years is the creation of a new class of security buyer: the indirect investor. The growth of pension funds and investment trusts has resulted in an "institutionalized" stock market. The foregoing introduces a new element of stability to the price level—but it also means that: (1) The dominant buying interest is confined to a relatively small portion of the available issues (2) A disparity in values has been created which eventually must be corrected. If the interest of the security buying public can be revived everyone will benefit—industry as well as the investor. The broader the demand, the more likely all stocks will reflect their basic values. An active stock market is synonymous with a vigorous economy.

Trends in Monetary Structures in 1954. M. Dutton Morehouse, *Brown Bros., Harriman & Co.*

In the nonbank long-term market, the supply of funds could well exceed the private demand by some $4 billion. In the short-term bank market the supply of funds will be governed by Federal Reserve policy, but there should be a negative net private demand of some $2½ billion. The Treasury should be able, therefore, to achieve measurable progress in extending the maturity of the national debt without running the risk of upsetting the market. The Federal Government might show a surplus of receipts over expenditures in the first half of 1954 in the magnitude of $7 billion and a deficit of $10 billion in the second half, or a net deficit for the calendar year of $3 billion. Administration efforts to reduce expenses might well result in a smaller deficit.

The decline in private demand for bank credit must be replaced by Government credit to avoid the deflationary pressure of a decrease in money supply. It is evident, therefore, that the bank market not only can but should absorb an amount of Government securities equivalent to the estimated deficit for the calendar year 1954. For the first time in many years the Treasury will find a market receptive to the offering of long bonds, and yet monetary policy appears to dictate the placing of a substantial amount of additional Treasury securities within the banking system. A ready market should be found within the banking system for intermediate maturities at yields equivalent to or somewhat lower than those prevailing today. Throughout the year the money market will be one in which a surplus supply of private funds seeking investment will be pushing towards lower yields. Frequent Treasury offerings will be designed to absorb these surplus funds as they accrue and thus prevent yields from changing substantially. Nevertheless, higher prices and lower yields are expected to prevail a year from now.

Monopoly, Bigness, and Progress. G. Warren Nutter, *Yale University.*

It has been often and widely argued that monopoly and big business are not only an inevitable consequence but also a necessary stimulant of economic progress. There is an element of truth in this proposition, but it emerges only after the argument has been severely qualified. In the first place, active competition can and does persist in industries characterized by continual growth and innovation Moreover, for the economy as a whole competition has scarcely diminished in vigor or extent in the face of the rapid and continual growth over the last half century; if anything, it has increased. In the second place, friction, in the form of delayed reactions to the lure of high profits, is the condition needed to induce innovations; and there is enough friction in most competitive industries to call forth

a constant stream of new products and technological improvements. Some of the outstanding new products of recent decades, like the automobile and radio, were introduced into competitive surroundings. Monopoly makes possible, though it does not ensure, extraordinary friction; and therefore it increases the likelihood that risky and costly innovations will be made. Large size combined with monopoly, a combination that is not inevitable, encourages the most risky and costly innovations. We now know very little quantitatively about the respective roles of competition, monopoly, and big business as promoters of economic progress. If discussion is to lead to intelligent changes in public policy, we must improve our understanding of these roles together with our understanding of how organizational forms affect other important values of our democratic and libertarian tradition.

Validation of Morbidity Survey Data by Comparison with Medical Records. NEDRA B. BELLOC, *California State Department of Public Health.*

A household sample survey of morbidity in the general population was conducted in San Jose by the California Department of Public Health in the spring of 1952. The survey was designed to test several methods of measuring morbidity and to study the completeness and accuracy of the information obtained. For validation purposes, medical records were collected from a number of sources. With respect to hospitalized illness, it was possible to study the net error in reporting. Admission rates based on household survey reports did not differ significantly from those based on hospital records for the same population. Similarly, days of hospitalization per person per year and average length of stay per period of hospitalization were measured accurately from household survey data. Distributions of admissions from household survey reports of hospitalization by month of admission, length of stay, and diagnosis were similar to the distributions obtained from hospital records. Whether or not surgery was performed was reported accurately in the household survey, but the description of the surgical procedure was not as precise as that obtained from hospital records. This study shows that reports of hospitalization obtained in a household sample survey are sufficiently accurate to be used for many purposes in lieu of hospital record data.

Hospital Morbidity Reporting—Experiences and Findings of a Pilot Project. I. Methodological Experiences. CARL L. ERHARDT, *New York City Department of Health.*

The Hospital Morbidity Reporting Project had two major aims: to test a method for collecting and tabulating data regarding patients discharged from hospitals in the city, and to analyze a meaningful body of the data collected. Both general and special hospitals, operated under municipal and other auspices, participated. Data covering the complete hospital system operated by the city are available for analysis. Reports submitted for each patient discharged during the project period included the usual demographic descriptions plus medical diagnoses, data on operative interventions and length of hospital stay. The facts, including diagnostic and operative data, were reported in code or in self-coding form, eliminating the usual expensive central processing step. The reported diagnostic code (Standard Nomenclature of Diseases and Operations) was mechanically translated into a statistical grouping of diagnoses (International Statistical Classification of Diseases, Injuries and Causes of Death) for analysis. The punching of digits of the Standard Nomenclature code was limited to those necessary for ISC classification. A condensation of the ISC was also designed for summary tabulation of hospital diagnoses significant in frequency in the New York City area. The methodological experiences and analytic procedures of this pilot project provide necessary background on which to base plans for routine collection of hospital morbidity data in New York City.

Hospital Morbidity Reporting—Experiences and Findings of a Pilot Project. II. Significance of Findings. MARTA FRAENKEL, *New York City Department of Hospitals.*

Medical diagnoses together with demographic data on hospital patients can contribute to morbidity statistics needed by public health administration. In conjunction with information on duration and outcome of hospital care, these morbidity data have significance for hospital administration and medical care planning. To date, most hospital statistics cover all patients disregarding their diagnoses. The analyses of the Project data showed that the composition of the patient load by age and diagnosis of 14 municipal general hospitals differed substantially. These differences explain many deviations in the data on "average length of stay," "net death rate," etc. if computed in the traditional unspecific way. It is shown how disease-specific data would permit more meaningful evaluation of hospital services and provide more useful bases for estimating personnel needs. Project data supplied disease-specific information on chronic illness. Many chronic disease patients stay only shortly in general hospitals. Short-stay hospitalizations—for special diagnostic, therapeutic or rehabilitative measures or for care during acute exacerbations—are elements of a long-term medical regime of many of these patients. Community-wide organization of integrated medical programs for chronic illness requires current disease-specific information on hospital care to these patients.

Statistical Determination of Tolerances in Rocket Development. EDWIN L. CROW, *U. S. Naval Ordnance Test Station.*

The paper considers the problem of determining the tolerances necessary for the components of a rocket in order that it have a prescribed dispersion. The establishment of the relation between departure from nominal construction and the resulting deviation of the rocket from the ideal trajectory is the main concern of this study, but two other requirements and their utilization are also discussed briefly: the probability distribution of the departures under manufacture within any given tolerances, and the relation between any particular tolerance and the cost of attaining it. To determine with satisfactory precision the relation between the deviation in trajectory and departure in construction of a particular rocket, an unusually large departure was purposely introduced into 50 rockets, which were then fired along with a control group. Firing was repeated with the departure doubled, and several types of departure were considered. In general the departure affected not only the mean coordinate of the trajectory but also the variances. A maximum likelihood estimate of the magnitude of the deviation was not practically attainable. An unbiased estimate of the mean square deviation was obtained by the method of moments, and its variance derived and investigated for various cases.

The Up-and-Down Method with Small Samples. J. L. HODGES, JR., *University of California (Berkeley)*

The main content of this paper is a report on computations made to determine the actual performance of some estimates for the mean dosage parameter μ, based on up-and-down series of length 10 or less. It is found that the Dixon-Mood formula for the asymptotic variance is reasonably reliable even in samples as small as 5 to 10. Thus, restriction (b) is not necessary. As a consequence of this fact, the design of up-and-down experiments may be altered, so that several independent series are run simultaneously, without serious loss of accuracy. In this way, restriction (a) can be considerably reduced. Finally, the possibility of conducting an experiment with several short series run in parallel introduces a flexibility into the design, which permits us to take advantage of special features of some experimental situations with a still further increase in efficiency.

Use of Statistics in Engineering in France. ANDRÉ G. LAURENT, *University of Chicago.*

In France, statistical applications in engineering are chiefly concerned with quality control with stress on: a) teaching, evidenced by creation of a training center in statistical methods for engineers at the University of Paris; b) publications, evidenced by books by Dumas, Laurent, Mothes, and a periodical, *Revue de Statistique Appliquée*, c) applications to industry, evidenced by applications in mining, mechanical, metallurgical, and textile industries. Difficulties arise from lack of financial support, poor reputation of statistics among industrialists, and the gap between theory and practice. Further developments may be expected along several directions: a) widening the fields of application; b) shifts of emphasis to descriptive and praxeological aspects as well as upon inference; c) more widespread use of available but little applied methods.

An Example of a Fractional Replication in a Bearing Abrasive Wear Test. F. R. DEL PRIORE and WILLIAM J. KOMMERS.

The purpose of this study was to develop an accelerated test for determining the relative wear resistances of plastic bearing materials. First it was necessary to find the effect of the test conditions on bearing wear. Physical considerations and limitations played the usual important role in the testing schedule. Six variables, each at two levels were investigated as a quarter-fractional. Three bearing materials were tested as split plots. The details in setting up the test plan are given as well as the analysis of the data. Following the analysis of the first quarter-replicate, it was considered that adequate results were obtained to obviate the testing of additional quarters. A special bench testing machine was designed and built on the basis of this work. Fixed test conditions were selected that produced the maximum differentiation between the wear resistances of the materials in a given period of time. The machine makes it possible to evaluate bearing materials in a matter of hours instead of the former month-long simulated service test procedures.

Gasoline Mileage in Winter Day-to-Day Use. DAVID FRAZIER and RAY DECKER.

Research executives of a petroleum company suggested that mileages of motor gasolines in northern winter "home to work" driving might differ importantly because of differences in starting losses, or in warm-up and misfiring characteristics. Such differences, if they really existed and were of important size, might suggest how the company could gain a competitive advantage. Therefore mileages of 15 gasolines, 7 commercial and 8 experimental, were compared in a complex statistical experiment on 15 employee owned and driven cars during the winter driving season. Each car in the course of the experiment ran 7 of the 15 gasolines each for one week, the gasolines being assigned to cars according to a statistical

plan called a "15 by 7 Youden square." All cars were driven normally, and gasoline consumption recorded for each gasoline on each car was the difference between gasoline added and gasoline drained at the end of the week.

When the experiment was complete statistical analysis, which eliminated all inherent "car to car" differences in gasoline mileage and much of the "week to week" variation within each car, revealed no significant differences among the gasolines. Further, the 8 experimental gasolines were so compounded that they constituted a 2^3 factorial which, when analyzed, showed with improved power that variation of vapor pressure, olefin content, or polyolefin content within reasonable limits had no significant effects on winter gasoline mileage. Inspection of results showed that real differences among the gasolines large enough to be important commercially would almost certainly have been reported as significant in this experiment, and therefore it was concluded that gasoline composition within reasonable limits has no commercially important effect on winter gasoline mileage.

Operations Research as a Science. JOHN B. LATHROP, *Arthur D. Little, Inc., Cambridge, Mass.*

During the last few years, the term "operations research" has been applied to a growing body of investigative activity, directed at problems of decision and management in many areas of business, industry, and government. It has been described as the application of the scientific method in new areas of decision, with the broadening of the range of such application of such magnitude as to require the appellation of a new branch of science. Operations research is certainly closely related to other fields of science and engineering, such as systems engineering, industrial engineering, market research, etc., and is often concerned with problems of interest to statisticians. These relationships can be illustrated by pointing out some of the general characteristics typical of operations research, and the role played by statistics in some examples.

As in science, the primary objective of operations research is to understand, not to act. When an operation is understood, the action required to improve the process is often fairly evident. The scientific method, by a combination of quantitative hypothesis, observation, and controlled experiment provides this understanding by means of relatively simple models to describe complex situations. Rather than reasoning from the facts to the mechanism, as in some forms of statistical analysis, the operations research scientist sets up assumed models in order to deduce phenomena to check against observed facts. An excellent and well known example of this method was Newton's explanation of the apparently unrelated phenomena of planetary motion and objects falling on the earth, by the simple unifying concept of gravity. In a study to reveal the causes of corrosion in paper mill equipment, the number of possible factors—chemical, physical and metallurgical—was very large, as was the amount of observed data from what amounted to an uncontrolled experiment. Although variance analysis was useful in determining the relative importance of various classifications of corrosion, statistical analysis alone would reveal no causes. It was necessary to develop a model of corrosion from first principles, using the physical sciences, in order to obtain a useful understanding of the situation.

Other models have revealed the relationships between sales promotion effort and sales, between production schedules and production costs, between warehouse stocking procedures and service to customers, between salesmen's compensation plans and profits, and so on. In some, statistical methods played an important part, in others no part at all. Information theory, queuing theory, servomechanism theory, symbolic logic or simple punched card reproductions of an operation have all provided the basis for operations research models. To summarize, operations research is a branch of scientific research concerned more with reaching an understanding of an operation than with the relations among the numbers describing the operation.

Effect of Current Operating Experience on the Realization of Investment Plans. JEAN BRONFENBRENNER CROCKETT.

Comparison of the planned with the actual fixed capital outlays of manufacturing firms, as reported to the Department of Commerce and the Securities and Exchange Commission, has indicated that individual firms' deviations from annual investment programs are subject to large variance. The hypothesis is tested that some significant part of the variance can be explained by concurrent fluctuations in the sales, profits, and liquidity of the investing firms. These variables may influence investment decisions either by affecting the expected return from proposed investment (in the first two cases) or by affecting the ability to finance desired expenditures (in the third case). The data studied cover about 400 firms in the mild recession year of 1949. Information on actual and anticipated investment and sales was supplemented by income and balance sheet data relating to profits and liquidity.

The extent and nature of the effect on investment decisions of the explanatory variables may be expected to depend on whether the proposed investment is primarily for expansion or primarily for modernization. Direct information was not available, but a rough attempt was made to separate firms which might be, from those which probably were not, interested in expansionary investment. For the

first group, deviations from investment programs were found to be correlated positively and very significantly with deviations of actual from anticipated sales, while for the second group the effect of sales deviations was not significant.

Changes in profit rates, as compared with the previous year, were also found to have a significant effect on the investment deviations of "expansionary" firms, but this was less marked than the influence of sales deviations and impossible to separate from it because of high correlation between the two explanatory variables. For the "non-expansionary" group, the effect of profit movements was not significant, and (except for cases of very large profit declines) the relationship appeared to be negative, possibly indicating a tendency for modernization expenditures to vary inversely with profit movements arising from changes in labor and materials costs. The effect of changes in liquidity on investment deviations is difficult to ascertain. However, in those cases where the observed liquidity movement is clearly something more than a *result* of the deviation from investment program, it appears to be a significant factor in *motivating* the deviation.

Expectations, Plans and Capital Expenditures: A Synthesis of Ex Post and Ex Ante Data. ROBERT EISNER, *Northwestern University.*

Ex ante and *ex post* data relating to sales and capacity changes, profits, capital expenditures and other variables have been derived from replies in McGraw-Hill questionnaire surveys and from independently obtained corporate balance sheet and income statement information. Analysis is focussed on the determinants of investment and, in particular, on evidence of operation of the acceleration principle. An "acceleration component" of investment is isolated and observed. Its absolute magnitude is found to be positively related to both current and lagged actual changes in sales and to expected changes in sales. And a larger "acceleration component" is associated with a larger total of investment. This evidence seems to fit well assumptions of recent economic theory of Hicks and Harrod which suggests a great significance for even a multilagged and quantitatively limited role of the acceleration principle. Investment is also found to be positively related to the rate of profit.

Expectations of sales increases are related to very great or long-term growth, but current sales changes are generally related negatively or not at all to expectations of immediately future changes. The directions of expected changes, but not their magnitude, can be matched to corresponding actual changes in sales. Discrepancies between planned and realized investment are found to be a decreasing function of size of firm, with very large firms appearing slower to adjust in a period of rising expenditures. Sales changes different from those expected, particularly in the case of firms expecting increases, are found to be a factor in the difference between investment plans and realizations.

Teaching Statistics to Executives. W. ALLEN WALLIS.

My discussion is based on experience in teaching statistics in the Executive Program at the School of Business at the University of Chicago. The Program's students are experienced executives. The objectives of the course are to teach: (1) avoidance of common logical mistakes in reasoning from data; (2) techniques for organizing numerical data into comprehensible form; (3) appreciation of the problems of allowing for chance and for unanticipated effects in designing investigations and in drawing conclusions from them; (4) a few rough-and-ready methods of analysis; (5) an appreciation of the potential uses of sampling; (6) recognition of a statistical problem when encountered; and (7) respect for the intellectual discipline underlying modern statistics.

Although the course operates mainly through lectures, the atmosphere is informal; class members raise questions and make many contributions, and time is left for post-lecture discussion. Numerous realistic examples are used, not only from business but from other fields. While we try to bring out the general principle illustrated by an example, a systematic exposition of statistical principles would have fairly low priority. What the students want, and what we are trying to give them, is something which will enable them to understand and generalize the experiences they have had so they may better cope with the broader experiences they expect to have. An important teaching device is the term paper. The students submit written descriptions of a statistical problem (preferably connected with their own business) and of the methods they propose to use for attacking it. This is returned to them with comments and suggestions, and must be approved by about the middle of the term. The final papers, submitted at the end of the term, are read carefully, and many students receive letters from the instructor commenting on them in detail. A statistics graduate student attends the lectures and writes detailed notes for distribution not later than the next class meeting. Regular textbooks are also used, and some exercises and quizzes are given.

We have come to feel that the kind of course developed for the Executive Program is more suitable on the campus than the kind of introductory course traditionally given. Consequently the courses in the Business School and in various other departments are being unified along the lines developed in the Executive Program.

Intercensal Needs for Small Area Data—By a Local Planning Agency. HARLIN G. LOOMER, *Philadelphia City Planning Commission.*

The statistical information that is needed by a local planning agency during the intercensal periods may be divided into two classes: that which is available or can be developed from local sources, and that which requires the special resources and authorities of a federal agency for collection and development. The former should be fully exploited before federal agencies are called upon for assistance. In the present situation of curtailed budgets, federal agencies should not undertake a multiplicity of new field survey projects, even on a sample basis, but should concentrate on fully exploiting and coordinating material that already is being collected by several such agencies. If all of the usual complete enumerations of the Census and all of the continuous records that are now being maintained by such agencies as B.L.S., Federal Reserve Bank, Old Age Assistance, etc. were properly brought together and integrated, there would be sufficient indices to indicate the trends from one decade to the next. Federal-local cooperative research should be further developed, possibly through the expansion of the facilities of regional or local branch offices of the federal agencies. Trained and experienced personnel from these offices could render valuable service in planning and directing local research. As a permanent continuing agency, these offices could bridge the gaps between intermittent or one-time surveys by local agencies and coordinate all into a comprehensive whole for the area. The authority and reputation of a federal agency (such as the Census) is needed to assure the success of some types of studies. Costs should be shared by participating agencies. Uniform sample surveys are of little value to local planners. Either the sample is too small to indicate purely local trends or the procedure is too standardized to meet specific local needs. A highly flexible cooperative arrangement would best serve the needs of local planning agencies.

Business Uses of Intercensal Data. W. R. SIMMONS.

Excepting local establishments, it is safe to venture that there are few business firms which do not make many vital planning and operational decisions based directly or indirectly on the findings of the Census. Generally taken for granted, like the mail and telephone service, Census reports become conspicuous only when and if they are absent. Not the least of the considerable American economic progress has been due to more intelligent planning based on sound factual information, to which Census reports are indispensable. In establishing plant and store locations, in developing local, regional, and national sales potentials, in defining sales territories, in analyzing problems of distribution, in evaluating sales performance and in deciding where, when and how to improve future performance, this type of information becomes crucial. Any substantial reduction in this service, in the commendable interest of economy, must run the risk of impairing the efficiency of our entire economy.

Meeting the Needs for Small Area Intercensal Data. CONRAD TAEUBER.

The major method of meeting needs for intercensal data for small areas from Federal sources is through Special Censuses conducted at the request and expense of local areas. Nearly 200 such Censuses have already been conducted since 1950, primarily for the purpose of establishing base figures for the allotment of tax funds. Although considerable interest in a quinquennial census of population about 1955 has been expressed, the prospects for it do not seem bright because of the costs. Estimates of population change since the last census and projections into the future are prepared locally by many cities and counties. The Census Bureau has issued a report suggesting methods of estimating current populations of small areas. School censuses are well established. but generally are not conducted in a manner to provide general purpose statistics. They represent a major resource for local area data which needs further development. There is need also for further research into the utility of the wide variety of local sources which could be used as indicators of population change in individual localities.

Industrial Production Index. CLAYTON GEHMAN, LORMAN TRUEBLOOD, ARTHUR L. BROIDA, M. H. SCHWARTZ, MILTON MOSS, PETER M. CODY, *Federal Reserve Board.*

The Board of Governors of the Federal Reserve System on December 18 released the basic postwar revision of the index of industrial production. The more obvious major features of the revision are an up-dating of the comparison base period from 1935–39 to 1947–49 and a change to the Standard Industrial Classification from the prewar Census classification system. More substantive features relate to the up-dating of the weight year from 1937 to 1947, enlargement of the number of monthly series from 100 to 175, substantial improvement of many old series, the development of independent annual indexes to which the monthly indexes are adjusted, and new seasonal factors for all major groups.

The revision so far applies mainly to the period from 1947 to date. For the earlier period the Board has adjusted its old manufacturing output indexes for the years 1939–1947 to the comprehensive benchmark changes developed by a joint Census-Federal Reserve project, and has linked the old to the new in January 1939 for the period back to 1919. Similar benchmarks and links were made for minerals

These interim revisions have been done at the level of the major divisions of the index in order to facilitate comparisons with the more recent period. In a general way, changes in industrial activity since 1947 are shown to be similar by both the new and the old total indexes. In the first half of this year, both indexes indicate that activity was at a record level for the postwar period, about one-eighth above a year ago. Both show that since midyear, output has been reduced fairly generally. For October, the new index was about 4 per cent below the highs which it established in May and July; the old index was down about 5 per cent from its peak reached in March. Both indexes show that industrial production in October was at about the same level as a year earlier.

Demonstration Teaching of Statistics. R. E. Wagenhals.

You can interest people in the use of statistical quality control without the use of mathematics by dramatic factual demonstrations. With the use of a series of demonstration devices, people can be shown how ineffective 100% inspection and 10% sampling may be as compared to scientific or statistical sampling. The repeated winner of a modified game of dice demonstrates he is not lucky but is just using control chart principles. A small pin ball machine demonstrates how a control chart and the normal distribution is related. The advantage of having quality controlled products is demonstrated by five different colored groups of twenty tubes to show how assembly tolerances can be reduced. These devices are demonstrated and the source of information regarding their construction will be available.

Use of Experiments in Engineering Statistics. Irving W. Burr, *Purdue University.*

The paper presents convenient populations for measurements and binomial and Poisson distributions. Then experiments to illustrate the following are discussed: sample frequency distribution, control charts, significance tests, estimation, analysis of variance, statistics of combinations, acceptance sampling, sequential analysis, linear correlation, and curve-fitting. A simple method of simulating a normal bivariate population is presented.

Using the Experimental Approach in the Teaching of Statistics. Edwin G. Olds, *Carnegie Institute of Technology.*

After a brief discussion of types of experiments which might be used, the paper discusses the value of the experimental approach in connection with lectures by the instructor and laboratory work by the student. The importance of using adequate time for orientation is emphasized. The paper closes with a proposal that the opinions stated be put to test by means of a designed experiment.

Problems in Experimenting with the Application of Statistical Techniques in Auditing. John Neter, *Syracuse University.*

Sampling is used extensively in auditing. Before any attempts to experiment with the use of statistical techniques in auditing can be made, one must formulate the purpose of sampling in auditing in such terms that statistical models can be constructed either for the problem as a whole or for definite parts of it. Once this has been done, the relative merits of various statistical models can then be investigated. One basic approach to the purpose of sampling in auditing would be to ascertain whether the figures in the financial statements are reasonably close to the figures which would be obtained if a complete examination of all transactions and records, including any necessary corrections, were made. Another approach would place emphasis upon an examination of the basic company accounting policies and of the effectiveness of the internal control process in assuring the accurate recording of transactions in accordance with these policies. Still another approach would encompass both of these points of view. Once the purpose of sampling in auditing has been suitably formulated, problems such as the determination of the alternative decisions from which a choice is to be made, the criteria according to which the appropriate decision is to be chosen, the relevant statistical model for the population under study, and the most appropriate sampling procedure can be studied and appropriate experiments conducted to determine the suitability of various statistical approaches.

A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. Robert Bechhofer, *Cornell University.*

A Two-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with a Common Unknown Variance. Charles W. Dunnett, *Lederle Laboratories.*

A Sequential Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. Milton Sobel, *Cornell University.*

A multiple decision approach is considered to the problem of partially or completely ranking a given set of population means. The particular goal of finding the largest of k population means is treated in detail. Let $\delta_{k,k-1}$ denote the true, unknown difference between the largest and second largest population means. It is assumed that the experimenter can specify before experimentation starts: (1) The smallest

value, say δ^*, of $\delta_{k,k-1}$ that he is interested in detecting and (2) The smallest acceptable value, say P, of the probability of achieving this goal when $\delta_{k,k-1} \geq \delta^*$. Three multiple decision procedures to accomplish this goal are given: (1) A single-sample procedure when the common variance is known. (2) A two-sample procedure when the common variance is unknown. (3) A sequential procedure when the common variance is known The observations X_{ij} are normally and independently distributed chance variables $N(X_{ij} \mid \mu_i, \sigma^2)$, ($i = 1, 2, \cdots, k$; $j = 1, 2, \cdots$ ad. inf.). We assume that the μ_i are unknown. The ranked μ_i are denoted by $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$. It is not known which population is associated with $\mu_{[i]}$ ($i = 1, 2, \cdots, k$). We define $\delta_{ij} = \mu_{[i]} - \mu_{[j]}$ ($i, j = 1, 2, \cdots, k$).

Single-Sample Procedure. (1) Enter the appropriate table with k and the specified P, and obtain a constant $h = h(k, P)$. (2) Set h equal to $\sqrt{N}\lambda$ where $\lambda = \delta^*/\sigma$ and σ is the *known* population standard deviation. Solve this equation for N. (3) Take N observations from each population where N is the smallest integer greater than or equal to the solution obtained in (2). (4) Calculate the k sample sums. (5) Select the population which yielded the largest sample sum as the population having the largest population mean.

Two-Sample Procedure. (1) Take a first sample of N_0 observations from each of the k populations. Any integer N_0 will satisfy the requirements of the problem. (2) Calculate

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{N_0}\left(X_{ij} - \frac{1}{N_0}\sum_{j=1}^{N_0} X_{ij}\right)^2$$

which is an unbiased estimate of σ^2 having

$$n = k(N_0 - 1)$$

degrees of freedom. (3) Enter the appropriate table with $n = k(N_0 - 1)$, and the specified P, and obtain a constant $h = h(n, k, P)$. (4) Take a second sample of $N - N_0$ observations from each of the k populations where N is the smallest integer equal to or greater than the larger of N_0 and $2s_0^2(h/\delta^*)^2$. (5) Calculate the k over-all sample sums $Y_{iN} = \sum_{j=1}^{N} X_{ij}$ ($i = 1, 2, \cdots, k$). (6) Select the population which yielded the largest over-all sample sum as the population having the largest population mean.

Sequential Procedure. Let the ranked sums based on m observations be denoted by

$$Y_{[1]m} < Y_{[2]m} < \cdots < Y_{[k]m}$$

and let

$$D_{im} = Y_{[k]m} - Y_{[i]m} \qquad (i = 1, 2, \cdots, k - 1)$$

"At the mth stage of experimentation take an observation from each of the k populations and compute

$$W_m = e^{-\delta^* D_{1m}/\sigma^2} + e^{-\delta^* D_{2m}/\sigma^2} + \cdots + e^{-\delta^* D_{k-1,m}/\sigma^2}.$$

(a) If $W_m \leq (1 - P)/P$, stop experimentation and choose the population which yielded the largest sum, $Y_{[k]m}$, as the one having the largest population mean. (b) If $W_m > [(1 - P)/P]$, take another observation from each of the k populations and compute W_{m+1}. Continue in this manner until the rule calls for stopping."

An Optimum Slippage Test for the Variance of K Normal Populations. DONALD TRUAX, *Stanford University.*

A solution is given to the problem of deciding if several normal populations with unknown means and variances have equal variances, or if not, which population has the largest variance. Under some mild restrictions on the class of possible decision procedures this solution is shown to be optimum in the sense that it maximizes the probability of making the correct decision when all the variances are equal except one which is larger than the rest. The optimum procedure is given as follows: select population Π_M if $s_M^2/\sum_{i=1}^{k} s_i^2 > L_\alpha$ and decide that all the variances are equal if $s_M^2/\sum_{i=1}^{k} s_i^2 \leq L_\alpha$. We denote by s_i^2 the unbiased estimate of the variance from the ith population, and M represents the index of the population having the largest sample variance. The constant L_α is determined by the restriction that if all the variances are equal then that decision should be made with probability $1 - \alpha$.

The Use of Statistical Techniques in the Aging of Accounts Receivable. R. M. CYERT, *Carnegie Institute of Technology.*

The aging of accounts receivable in most enterprises is done by sampling a portion of the accounts receivable. On the basis of the aging an allowance for uncollectibles is estimated. The study described in this paper had two primary objectives: (1) To analyze the sample design being used for selecting accounts to be aged with a view toward improving the method of selection and determining the optimum sample size. (2) To examine the plan for selecting accounts already aged to be test-checked by the auditor

with a view toward installing an appropriate acceptance sampling scheme. The procedure followed was to examine a past sample drawn from a large metropolitan department store. The sample was first analyzed for randomness by the use of Kendall's coefficient of rank correlation. There was no significant correlation found between the alphabetical rank of the accounts and the rank of the account age. It was decided, therefore, that it would be meaningful to compute the precision with which the past sample estimated the allowance for bad debts. The precision as computed from the sample of 15,000 accounts was deemed too tight by the public accountants working on the study. New precision and reliability requirements were specified and a new sample size was computed on the assumption of an unrestricted random sample design. The required sample size was reduced from 15,000 to 1,700. The sample design specifies that the accounts will be drawn systematically and the sampling error will be computed by the Tukey plan.

Previous practice in test-checking the agings done by store personnel had been to select 10% of the accounts aged. The public accountant then decided whether or not to accept the total accounts aged on the basis of the test-check. The process of test-checking was one to which sequential analysis was applicable. After the accountant had defined a good and a bad lot and specified the risks to be taken, a sequential sampling plan was determined. The plan was truncated at three times the average sample number for lots with a per cent defective equal to the slope of the decision lines. The maximum number that would need to be sampled, therefore, is 376 and the average sample number is 126. In the authors' view, this study establishes the possibility of adapting statistical methods to accounting and auditing problems involving a high volume of homogeneous data. It is important to note that the subject of this study and, therefore, its conclusions do not relate to an area of audit in which results are applied directly to the decision of giving or withholding an opinion on financial statements.

New Experimental Designs for Paired Observations. W. S. Connor, *National Bureau of Standards.*

There are many experimental situations in which only two objects can be observed at a time under homogeneous conditions, as for example in a recent experiment at the National Bureau of Standards which involved the comparison of several types of spark plugs in two cylinder gasoline engines. In these cases experimental designs with two plots per block are needed. If all possible pairs of objects are observed, the statistical analysis of the observations is simple. However, if the number of pairs is large, this procedure may be expensive and may provide more precise results than are needed. Thus it often is desirable to observe only part of the pairs. The choice of this subset of pairs must be made with care if the statistical analysis is to remain simple. Such a subset is the "two-group arrangement," which was suggested by the traditional experimental procedure of comparing several new objects with one or more standards. The two-group arrangement consists of dividing v objects or treatments into two groups of m and n treatments ($v=m+n$), and of pairing every treatment from one group with every treatment from the other group. No other pairs are formed. The analysis appropriate to this arrangement is described in detail.

Some Aspects of Sequential Experimentation. S. Munro.

Out of practical necessity a sizable amount of industrial and military development is being conducted sequentially, even to the extreme of a sequence of samples of one. There is, however, aside from a handful of exploratory and superficial papers, no body of mathematical theory which serves as a basis for the design of such experimentation. Since this experimentation is performed to make use of all the available information, in the colloquial sense, at every stage in the sequence, the usual statistical hypothesis of independence is not valid. Also, unlike classical design problems, this experimentation is devoted to estimation rather than tests of hypotheses. Many different types of practical problems will be discussed and an attempt will be made to show that the transition from one experiment to the next is a statistical problem, properly subject to mathematical study.

Information Theory and Prediction. Max. A. Woodbury.

A study is made of the usefulness of information theory as an aid in making and evaluating predictions. Clearly, predictions are made on the basis of information available to the predictors. Two principal problems are faced in making a prediction: First, the selection of the prediction, and second, evaluating the prediction in the light of what happens. Note that a prediction involves a selection of information available and a transformation. Naturally there is interaction between the two problems since one would change the method of prediction appropriately if the verification methods change. Here we use the measure of information proposed by Shannon and Wiener. A value is attached to probability distributions similar to entropy or negative entropy and the amount of information is measured by the change in entropy due to the additional information. In practice this involves the use of Bayes theorem. The principal predictions studied in the paper are meteorological forecasts. An ONR project has been set up to evaluate the methods proposed and with the help of the U.S. Weather Bureau

and the U.S.A.F. certain studies have been made. Here the verification evaluation consists of computing an "information ratio." This is the amount of information available in the forecast relevant to the phenomenon predicted divided by the amount necessary for a "perfect" forecast. Techniques for discarding irrelevant information have been investigated and the use of discriminant and factor analytic methods have been undertaken.

Estimation of Length of Hospital Stay from Discharge Data. CLIFFORD A. BACHRACH, *The Johns Hopkins Hospital.*

This paper considers the relationship between the two distributions of patients' length-of-hospital-stay that are commonly tabulated and used in hospital administration. These are:

1. The distribution of the lengths of stay in the hospital of patients who are discharged during a time interval.

2. The distribution of lengths of stay-to-date, of patients who are in the hospital at a given time. The event leading to a patient's inclusion in a group of patients discharged during an interval is discharge—an event which happens only once to each patient. The event leading to inclusion in a sample of the second type is presence in the hospital on a particular day—an event which may occur repeatedly to a given patient, with a frequency depending on total length-of-stay. Hence, patients who are in the hospital at a given time may be regarded as a length-biased sample of the hospital's patients; the expected mean value of the stay-to-date of such a sample is given by $(\sigma^2/2\mu)+(\mu/2)$, where μ and σ are population parameters of the distribution of lengths of stay of patients discharged during a time interval.

An analogy is drawn between a general population and a hospital patient population. The distribution of stays of patients discharged during a time interval is the analogue of the distribution of ages at death of a general population. The distribution of stays-to-date of patients in the hospital at a time is analogous to the age distribution of the general population at a census. If the stationary population model fits a hospital population, the two length-of-stay distributions are related in the same way as the d_x and l_x curves of a life table. The assumptions underlying the stationary population model are considered with reference to hospital populations of various kinds. The effects upon the model of such events as change in bed capacity and time trends in length-of-stay are pointed out. The assumptions are such that they are more likely to be satisfied in a hospital with short-term patients than in a chronic disease hospital. Data are presented to show that in one large civilian general hospital (The Johns Hopkins Hospital) the stationary population assumptions are sufficiently well satisfied to permit the use of this model for the purpose of estimating the characteristics of the distribution of the length-of-stay to-date from data on discharged patients.

Parametric Estimation of Survivorship. D. J. DAVIS, *The Rand Corporation.*

This paper deals with the application of statistical methods to the development of failure distribution hypotheses for complex mechanical systems and to the testing of these hypotheses. The two failure hypotheses considered are: *Exponential* in which non-failed systems are assumed to exhibit an equal probability of failure during equal time intervals independent of the previous operating time. *Normal* in which the probability of failure per unit time for unfailed systems increases with use in such a manner that the times-to-failures tend to cluster around the average value in a distribution similar to the normal. Statistical analysis is applied to failure data for a variety of systems involving people and/or electrical, electronic, and mechanical equipment. The raw failure data are arranged in frequency distributions which are compared to hypothetical distributions derived from either an exponential or a normal theory of failure, whichever is appropriate. The fit between the observed and hypothetical distribution is measured by means of the chi-square test. The chi-square tests indicate no significant disagreement between theory and observation in all but one of the 34 data samples examined. The 34 failure distributions are dominantly exponential rather than normal.

The Demand for Citrus Products. GEORGE M. KUZNETS, *University of California*, and RICHARD J. FOOTE, *Agricultural Marketing Service.*

The paper deals with the problems of utilizing the information collected weekly by the Market Research Corporation of America from a national panel of consumers on purchases and prices paid for various citrus items and canned juices to obtain estimates of demand elasticities. The advantages of utilizing data which exhibit variation both over households and over time are first discussed with reference to a simple model of purchase behavior. A comprehensive study is then outlined which involves separate regression analyses for summer and winter seasons extending over a recent period of 38 months in each of 10 rural-urban geographic strata. The unit of observation is a family-month and the variables include monthly purchases by panel families of each of six citrus products, prices paid for these and for four other products, family income, family size, an index of availability of frozen orange juice, and month within season. The main problem encountered in proceeding with the study is the absence of in-

formation on prices confronting the non-buying families. Probably not less than 70 per cent of the required prices for each product will have to be estimated each month. Fragmentary results from several pilot studies in progress are described These studies are designed in part to test the usability of panel data and in part to provide some indications on the basis of which a reasonable choice of a procedure of estimating missing prices could be made.

An Analysis of the Demand for Meat. AYERS BRINSER, HARRY ALLISON, and CHARLES ZWICK, *Harvard University.*

The purpose of this study of the demand for meat, fish and poultry meat items was twofold. In the first instance, it was an attempt to define certain factors affecting demand in addition to price and income and to determine the order of their relevance. The other factors include family composition, education, age, time available for meal preparation by the housewife, and ethnic group. The second objective was to test the effectiveness of a consumer panel to supply the data for such analysis. This paper is a progress report of the study outlining the design of the research and the method of data collection with a preliminary analysis of the consumer panel technique as it was applied. A full evaluation of this aspect to the study must wait until all of the data collected have been analyzed. The results of the study up to this point suggests that the consumer panel technique is a workable method for collecting data at the household level with limited research resources. It remains for the continuing analysis to show whether these data can be used to provide effective answers to such questions as whether there are differential price and income elasticities among the various subgroups in a population. On the evidence of preliminary estimates it would seem that the form and substance of the data collected will support useful quantitative analysis.

Demand Analysis from the M.S.C. Consumer Panel. G. G. QUACKENBUSH, *Michigan State College*

The M.S.C. Consumer Panel consists of about 250 families who report each week in considerable detail on their food purchases. Each family reports the quantity, price, and expenditure for each food item purchased. About 500 food items are included, and this includes all food. Each family reports its income for the week. The sample area to date has been the city of Lansing, Michigan. The project was designed to continue for 10 years. The objectives of the project are to determine price, income, and cross elasticities, as well as purchase patterns for food. The research was undertaken to help fill the need for cross-sectional analyses over time and for analysis of data of the short-period type. Examples of current research include use of single equation least squares methods to derive price elasticities for beef, pork, poultry, fish, cold cuts, individual cuts of meat (steaks, roast, hamburger, etc.), and some fats and oils. Cross-sectional analysis, using total annual purchases for individual families, has been used to derive income expenditure elasticities for food eaten away from home, food purchased for home preparation, and for some individual food products. It is concluded that the consumer purchase panel method has considerable potential as a method of securing data for demand analysis and related studies.

The Flow of Net Cash Savings through Life Insurance Companies. HARRIS LOEWY.

The flow of net cash savings through life insurance companies is made up of the savings of policy holders (internal savings) and governments, corporate debtors, and mortgagees (external savings) Internal savings equals companies' net income, and is the net of income flows such as first year's and renewal premiums, interest income, and expenditures flows such as payments to policyholders and expenses. The growing stability of internal savings is the result of the increasing relative importance of such contractual income flows as renewal premiums and interest income. Payments to policyholders will also stabilize as the result of the rapid growth of group and term insurance in force in which policyholders' equity is small. As equity declines in the future, policyholders' dissaving, a feature of the Great Depression, will also be relatively less.

Individual decisions to save are mainly incorporated in the first year's premium flow which shifts rapidly with changes in the level of income and employment. The renewal premium and interest income flows are relatively insensitive to income changes and are increasing in stability because of the contractual element they incorporate. Total income is stabilizing because of this contractual element and because the volatile first year's premium flow is becoming a declining proportion of total income. The expenditures flow is dominated by payments to policyholders. In the 30's the flow of payments to policyholders was strongly influenced by lapses and surrenders. If the decline in policyholders' equity continues as a result of the growth of term and group insurance, swings in this payments flow should be of smaller depth and amplitude. The net result of these developments should be growing stability in the annual flow of internal savings or net cash available to finance capital formation. The flows of external savings, amortization payments, are measured for 1947–1950. Added to internal savings, they show how much net cash (new money) the life insurance industry made available in those years by source to finance capital expansion.

Current Problems in Measuring Moneyflows. MORRIS A. COPELAND

Moneyflows measurements constitute a set of interlocking social accounts for the economy—sector sources and uses of funds statements. This system of social accounts is an extension of the national income and product system to show inflows and outflows for nine (instead of five) separate sectors, and to show a variety of types of transaction that do not appear in the national income and product accounts, particularly financial transactions. The way in which the economy must be divided into sectors for this purpose calls not only for replacing the business sector of the national income and product accounts with at least three financial and two nonfinancial business sectors; but several national income sectors must also be redefined. A critical current problem involved in the sectoring of the economy that is examined in some detail centers around the contrast between personal saving (disposable personal income minus personal consumption expenditure) and the net financial use of funds by households (increase in household cash and portfolios minus increase in household debts). The extent to which other regular statistical compilations (as well as the national income estimates) are currently used in making annual estimates of sector moneyflows is emphasized. Several problems involved in putting the various sources of information together into moneyflow accounts are outlined. An important but as yet unsolved problem is that of developing current quarterly estimates of moneyflows. The nature of this problem is considered in relation to the monthly and quarterly figures now available.

Estimation of the Interval Rate in Actuarial Calculations: A Critique of the Person-Years Concept. JOSEPH BERKSON, *Mayo Clinic.*

Each individual in a group for which a follow-up study is being made is observed from some defined origin of time, as for instance the time of receiving medical treatment. The observations may continue to some predesignated closing date C or be terminated by a random process. The problem is to estimate the probability of survival to time T_j. Broadly, two methods can be used. (1) If the probability of survival to T_j is given by some smooth survivorship function $P_j = F(T_j, \theta_1, \theta_2, \cdots)$ where $\theta_1, \theta_2 \cdots$ represent parameters, then with appropriate observations available, the θ's may be estimated by using some method of "fitting" the function, such as maximum likelihood, minimum X^2, least squares, etc. (2) If the function F is not known, the actuarial method is used. In this method the total period T_j is subdivided into intervals small enough so that within each interval the survivorship can be considered linear. The conditional probability of surviving to the end of the interval $[x, (x+1)]$, having survived to the beginning, is p_x, and if this is estimated for each interval, the probability of surviving to time T_j is given by $P_j = \prod_1^{x-1} p_x$. The essential problem, then, is to estimate p_x. Two estimators are discussed, (1) "Person-years" estimate, given by $\hat{q}_w = d/\text{P.Y.E}$, where d is the number of deaths observed in the interval and the "person-years exposed," $\text{P.Y.E.} = N_x - \sum_1^w (1 - T_{ix})$, where N_x is the number living at the beginning of the interval and T_{ix} is the fraction of an interval for which each of W "withdrawals" among the N_x was to be observed. (2) Maximum likelihood estimate, given by the solution of

$$q_L = d \Bigg/ \left[N_x - \sum_1^L \frac{(1 - T_{ix})}{1 - qT_{ix}} \right],$$

where the summation is taken over L individuals last observed living in the interval. Various approximations of each of the estimates are discussed and variances of the estimates are evaluated. It is pointed out that the "person-years" estimate implies that it is possible to define a "withdrawal" W. For the usual follow-up study a definition of W requires the imposition of a single "closing date" for all persons in the study. This involves the disadvantages that it (1) necessitates a truncation of the data that sacrifices available information, (2) makes clerical work more complicated than when the maximum likelihood estimate is used, (3) renders impossible calculation of survival rates for any series in which less than 100% of the cases are traced to the closing date.

The Problem of Within Family Contagion. WILLIAM R. GAFFEY, *University of California (Berkeley).*

The presence of contagion, or direct person-to-person transmission, in a disease process is often argued from the fact that the secondary attack rate is high compared with the primary attack rate. An example is given to show that no such inferences regarding contagion can be made from the comparison of these rates, because of the possibility of different intensities of exposure in different families. Instead, the observed variables are taken to be the times during the period of observation at which cases occur within a family, assuming families to be independent. A model is constructed which admits either (1) no contagion, (2) positive contagion (in which a case makes subsequent cases more likely), or (3) negative contagion (in which a case confers a degree of immunity on the rest of the family). The joint density of the times of occurrence referred to above is given under the hypothesis of no contagion and under the alternative of contagion (positive or negative) of a special type termed "linear."

A uniformly most powerful test of the hypothesis (1) against either (2) alone or (3) alone is found, as well as a uniformly most powerful unbiased test against both (2) and (3) together. All of these are insensitive to variations in the intensity of exposure. The first two moments of the test statistic were found as an aid to getting an approximate region of rejection when large numbers of families are involved.

Consistency of Estimators under a Specialized Bioassay Procedure. William F. Taylor.

The consistency properties of various types of estimates are considered in this paper, with especial emphasis on maximum likelihood and minimum chi-square estimates. A general theorem is presented stating necessary and sufficient conditions for consistency. It is assumed that the effects of a drug are to be evaluated by giving it in varying doses to several subjects and observing the number of these subjects who respond in a given way. Let the dosage levels be $x_1, x_2, \cdots x_s$ and the number of subjects receiving the ith dose, n_i. The proportion of subjects responding to the ith dose is called q_i. A further assumption is that the probability, p_i, of a subject responding to dose x_i is some function $f(\theta|x_i)=f_i(\theta)$, where θ is a parameter to be estimated. For several of the commonly chosen functions, f, and several methods of estimation, it has been shown that the estimates of θ so obtained are regular best asymptotically normal (RBAN) in the sense of Neyman, and thus have identical asymptotic properties.

There is more than one type of limiting situation which leads to asymptotic properties, however, and the fact that one such situation results in desirable properties may have little to do with the properties associated with other limiting situations. The results most frequently used are based on the case in which the total number of subjects is $N=\sum_{i=1}^{s} n_i$. N is allowed to increase by increasing each n_i, keeping the n_i/N and the s fixed. When this is the limiting situation considered, maximum likelihood, minimum χ^2, logit, and probit estimates are all RBAN When $N\to\infty$, not by increasing each n_i but by taking on more dosage levels, i.e., by increasing s, the above can no longer be said. This is shown with regard to the consistency of the estimates, assuming the n_i are all $\leq \bar{M} < \infty$. While maximum likelihood estimates can be shown, under certain conditions, to be consistent, under these same conditions minimum χ^2 ones are not consistent unless additional restrictions are imposed.

Use of Census Tracts in Study of Changing Residential Patterns in Metropolitan Areas. George Duggar.

For decision on issues of policy and theory research is needed on residential localization. Studies should be extended to encompass entire standard metropolitan areas and, in the search for a valid typology, applied with identical statistical techniques to several metropolitan areas and to several periods of residential construction. The 1950 Census of Housing permits comparison between standard metropolitan areas and between parts of areas. By comparison with the 1940 Census and by reference to data on year built new and older housing can be compared, and areas can be distinguished in terms of their rates of housing growth. Gross techniques for study of metropolitan areas are examined, centering on the 91 areas with a city of 100,000 or more and contrasting housing in their core cities and remaining area, and housing built 1940–50 and earlier. Correction of results for differences in extent of land area and use of the new urbanized area concept are found valuable. Techniques for more intensive comparative study of growth periods, of entire metropolitan areas, and of more central and more peripheral portions are examined, using census tracts and, in the untracted portions, other small statistical areas. Techniques include mapping of growth rates, growth concentration, and housing characteristics, statistical analysis of relations between these (in ways relevant to theory) and identification of small areas where growth rate or concentration suggests case study While lack of cross tabulations prevents direct comparisons between the housing in the several tracts by period built and characteristics, the general patterns of co-variation of growth rate and of housing characteristic can be examined as reflected in tract totals. It is suggested, however, that the type of distinction which has been drawn in the 1950 census between urbanized area and standard metropolitan area be drawn, also, between the urbanized and non-urbanized portions of tracts.

Theory of Behavior. Edward W. Barankin.

See *Econometrica*, Vol. 21, No. 3, July, 1953, p. 474.

Uses of Census Tracts by Housing People. Dorothy S. Montgomery.

Tract use gained impetus in Philadelphia as a result of interest in the two Real Property Surveys of the 1930's, as well as the 1930 Census. The Philadelphia Housing Association, the first housing group in the city, was among the first to use the tract for analysis and recording. As an educational, Community Chest agency, we find the tract a valuable tool in educational work as well as in our fact-finding activities. At the outset, the census tract helped to tell the story of the relationship of housing to

health, to delinquency, to dependency. Today our Association uses census data on a tract basis to interpret neighborhood housing conditions in connection with the housing tours which are an essential part of our educational program.

The technical uses of the tract are numerous, and most of them common to other communities. But I would like to describe two uses that may be unique. The first is a use that continues year after year as part of our tabulation and analysis of city permits covering new dwelling construction, conversions, and demolitions. These data have been regularly collected by our Association since 1923, and have enabled us to estimate changes in the dwelling supply throughout the city in intercensal periods with a reasonable degree of accuracy. Since 1945 these data have been tracted, which has added the important element of precise location to the previously tabulated information on type of new housing, selling price and lot size. The tracting of conversion data is of particular value since it has permitted analysis of the significant, and usually neglected changes, in the use of structures, changes that are frequently associated with the start of neighborhood blight. It may be mentioned here that the tracting of the permit data became feasible when an up-to-date street and house number directory was published for Philadelphia. Prior to the directory, tracting was almost impossible.

Tract data enabled us to discern and map the patterns in the spatial movements of Negro households in the city, and to gauge the amount of concentration and dispersion during the last census decade. Most of the influx of the decade had been absorbed in the traditional areas of Negro occupancy which had become more Negro in the process. Within these areas, however, there had been substantial shifts with the North Central Philadelphia area becoming the major center of concentration. Analysis of the tract data also showed that, contrary to general opinion, Negroes are not finding homes throughout the city; and that where expansion did take place, it was in or adjacent to areas in which Negroes already constituted a large part of the population. While over-all census figures indicated that housing occupied by Negroes was on the average of much poorer quality than housing occupied by white families, analysis of tract data revealed that there is no inevitable correlation between the race of the occupants and the quality of the housing. A comparison of two neighborhoods showed that while one, an area of high Negro concentration and severe overcrowding, has become increasingly blighted in the last two decades, the other, also an area of high Negro concentration but low density of population has shown an improvement in housing quality in the course of the racial change. This study, which could not have been undertaken without tract data, will provide the Commission on Human Relations with factual information upon which to base its educational program to end discrimination and segregation. It will also affect other aspects of the city's official policy concerning the enforcement of minimum housing laws.

Uses of Small Area Census Data in New York City. FLORENCE S. CUTTRELL. *Welfare and Health Council of New York City.*

The decade 1940–1950 brought so many shifts in the population of New York City that many organizations are studying the 1950 population and housing data for the 2563 census tracts, or the 352 larger health areas which are composed of tracts. One of the major uses by City departments, social agencies, churches and similar organizations is to estimate the needs of areas within the boroughs for housing projects, schools, health programs, case work services, recreation centers, programs to prevent delinquency, churches, etc. and to determine the location of facilities to provide these services. The City Planning Department utilizes small area data in connection with its master plan, zoning ordinances and its other responsibilities and for population forecasts. Several major research projects—studies of the aged, mental health, medical care and the teen-age narcotics problem—have used the data for selecting samples and/or for studying population characteristics and environmental factors. Utilities, commercial and savings banks, insurance companies, newspapers, construction firms and other companies use population and housing data to estimate loads or sales, to plan districts, for mortgage appraisals, to determine the need for housing or branch banks, etc. Manufacturers, advertising agencies, radio stations, and research organizations indicate dependence on block and tract data for sample marketing and public opinion surveys.

Special publications: To supplement the tract data published by the Bureau of the Census, six organizations purchased a set of the IBM summary tract cards on the 1950 population. These cards also enabled the City Department of Health and the Research Bureau of the Welfare and Health Council to publish population data for health areas. The cards also provided data for the Population and Housing volume of the "New York Market Analysis," a companion of the volume on retail trade from the 1947 Census of Business. Both volumes of this publication by the New York *Mirror*, New York *News* and New York *Times* cover 116 retail trade districts (aggregations of tracts) within the City and 21 suburban counties. The Research Bureau of the Welfare and Health Council published for census tracts and health areas: "Population of Puerto Rican Birth or Parentage, New York City: 1950."

Probability Distributions of Group Organization Theory. LEO KATZ, *Michigan State College.*

A functioning group of N individuals is represented abstractly as a multilinear, graded, directed graph on N points. Between each pair of points (individuals), the complete relationship is assumed to be analyzable into a sequence of categories, in each of which the strengths of the bonds in both directions are measurable in some scale. The hope for economy in description of the group rests on the possibility that the bulk of the information in these infinite-dimensioned vectors is contained in relatively few (perhaps, only one) of the components. A single component (one facet of the group organization), then, is represented as a directed graph and, if the relationship is all-or-none, as a simple linear directed graph. Two fundamental results provide an almost complete probability distribution theory for the classical sociometric problems. A thesis by James H. Powell of Michigan State College will contain an analysis of the structure of the sample space of sociometric investigation whidh makes possible the immediate, though not simple, expression of most of the pertinent probability distributions in terms of the numbers of graphs satisfying certain restrictions. A second result by Katz and Powell (submitted to *Proceedings of the American Mathematical Society*) gives explicitly the number of directed graphs satisfying a specific set of local restrictions

The principal classical problem not solvable as above is that of the distribution of mutual or reciprocated choices. However, an iterative procedure of application of the previous results produces the required distribution for this case. Applications of the general method include the measurement of concentration of choice, tendency toward reciprocation, development of leaders, etc. Finally, preliminary study of the bilinear graphs corresponding to two-dimensional group functions indicates that any attempt to handle these mathematically would involve invention of some new mathematics. On the other hand, this study indicates some new uses of multiple-level sociometric techniques in the field of social psychology.

Tracts in Analysis of Worker Mobility. MEREDITH B. GIVENS, *New York State Department of Labor.*

One of the uses of census tracts is to provide a basis for aggregation of data into special-purpose area groupings. Census tracts were recently utilized in this way by the research staff of the Division of Employment of the New York State Department of Labor, in a study of where people work and live in New York City and its metropolitan environs. The study was undertaken for administrative purposes, though its results are proving to be of considerable interest to all who are interested in worker mobility and the pattern of commuting in the metropolitan community. It involved comparisons, by small areas within the Metropolitan Area, of the place of work of employed persons as revealed by the Division's own employment data, compiled for employers of twelve or more workers, with the place of residence of workers living in these same areas as shown by the Federal Census of Population. Such comparisons of (1) employment at their work place with (2) Census data on employed persons counted at their homes are indicative of the "balance" or "imbalance" between the number of persons working within an area and the number of persons living in that area. The comparison does not show where the workers within a given area come from or where those who live within the area work. For this purpose a special question in the Census schedule would be required.

For purposes of analyzing and mapping data, the intra-State Metropolitan Area was divided into 57 districts. These districts represented combinations of census tracts in those localities which have been tracted, namely New York City and Westchester. For Rockland, Nassau and Suffolk, which are still untracted areas, other approaches were used for subdividing the counties. For New York City, the basic data were first prepared on a census tract basis and then combined into 210 "conversion districts"—a tentative grouping of census tracts developed with the assistance of a committee named by the local chapter of the American Statistical Association. This grouping was designed for possible use in obtaining tabulations of Census materials on a small area basis. Contiguous "conversion districts" with similarities in employment were then combined into 32 broader districts. For Westchester, the only other tracted county in the area, the 150 census tracts were combined into nine districts which conform to the developmental districts delineated by the Westchester County Department of Planning.

To obtain information on employment at place of work on a census tract basis, for combination into the broader districts, it was first necessary to code the data on insured employment according to census tract. To obtain employment data according to place of residence for combinations of census tracts, the Division purchased from the Bureau of the Census IBM summary cards showing, by tract, employment characteristics of the employed labor force for New York City and Westchester. The New York City purchase was made jointly with several other local groups using combinations of tract data for their own purposes. The task of coding the Division's employment records by tract proved difficult because of the lack of an up-to-date directory of census tracts for the City.

This project required the solicitation of one-time breakdowns of employment at each place of work from multi-establishment employers. The project demonstrates the feasibility of using census tracts as an approach to small-area analysis of data from other than Census sources in comparison with Census

data. The tract is a useful common denominator in devising larger standard areas for statistical tabulation and analysis in a metropolitan area.

Use of Census Tracts for Business Analysis. PERRY H. MYERS.

The main problem in the use of census tracts and other census data for business analysis is simply ignorance of even interested businessmen regarding census material. An informal survey suggests that about half of business executives think a census tract is a pamphlet issued by the Bureau of Census. Census tract data are essential in several major marketing developments: (1) In tracing the differentiation of the urban market and particularly the development of a new type suburban group. In this connection, it should be noted that higher incomes, the building boom, more marriages and more children have led to an accelerated movement of younger, middle income couples into the suburbs and, at the same time, concentration of lower income groups within the central parts of the larger cities. To trace this development, which has far reaching effects for both retailer and manufacturer, requires detailed study of tract data, not only in terms of population but also in terms of income, family characteristics, home ownership, etc. (2) The development of suburban retailing, particularly large department store plans for shopping centers, has required detailed analysis of local data to determine the population and characteristics of the areas to be served by these new suburban outlets. (3) Tract data are essential in consumer sampling, both in setting up a random sample and also in projecting the results to the total population of a city or to other cities. In general, it may be said that increasing purchasing power and the wider distribution of discretionary income have led to greater differentiation in the consumer market and a greater need of tracing this differentiation in terms of where people live, and who they are, through the use of the census tract.

The Relation of Census Tracts to the General Census Program. ROBERT W. BURGESS.

The activities of the Census Bureau, the methods used and the subject matters covered can be partitioned in various ways, many of them significant in directing the enterprise or proposing changes. One of the fundamental splits which effects both the determination of what shall be done and the way in which the details of various projects shall be developed is between the nation and the small area as the focus of interest. Typically an economist values statistics for the information they give of the nation as a whole and of the major economic divisions thereof. He is also interested in the changes over relatively long periods of time. The market researcher, however, and other specialists in planning economic and social activities, need information as to the population in relatively small local areas. They want to have an adequate basis for determining where a store or a school or a hospital can best be located and for judging whether one of these institutions is really meeting the present and future needs of its neighborhood. For such purposes sampling procedures do not fill the bill, and the small area statistics provided by a reasonably recent complete census are necessary.

The needs for small area statistics have thus an important bearing on determining the proper period between censuses covering a particular field. These needs also influence decisions as to the form in which census results should be stated, so as to meet the needs of users and encourage greater use of all material. It has become clear as workers in various fields apply scientific methods increasingly to the key factors of their problems that these needs for local statistics deserve careful attention. The Census Bureau, therefore, can and does cooperate with local agencies and individuals in developing census tracts and stimulating their use. The Census Bureau, however, leaves the functions of analyzing tract statistics and campaigning for educational or social projects that such analysis might suggest to individuals like subsequent speakers on this program or the organizations which they represent. Noteworthy gains are being made in the establishment of census tracts so that within a few months 46 standard metropolitan areas will be entirely divided into tracts as compared with only 11 at the time of the 1950 Census. With continued interest on the part of local committees, it is a reasonable goal that we have census tracts for all SMA's developed for the 1950 Population Census.

How the Automobile Industry Utilizes the Census Tract in Market Determination. FERDINAND F MAUSER, *Wayne University.*

The automobile manufacturers are fairly consistent in their belief that the census tract is a useful, basic statistical tool, of value in gathering distribution intelligence. Marketing strategy based upon this intelligence determines placement of dealerships. Companies consider proper placement of dealerships to be vital in their attempts to achieve maximum sales for their products. Two of the methods employed in area analysis using the census tract as the basic working framework are the standard metropolitan area approach and the central city approach. Superimposing of selected data upon area maps makes it possible to bring areas of poor dealer performance into focus. The type of data used for addition to the census tract framework varies depending upon the need and makeup on the area studied. There may be as many as 10 or 12 separate maps drawn up for a single area analysis. Statistical specifics are presented

on a three part basis: (1) the ideal situation—what the company would want if it could write the ticket exactly as it wanted, (2) the present situation as it now is, and (3) the recommended proposal in terms of specifics—the answer to the question of what should be done now. Under this system the ideal arrangement is used as the starting point. The situation as it now stands, is then related to the ideal. This brings the problems into focus. Changes then become evident and improvements are made gradually in the direction of the ideal because the ideal or goal is known.

Further Generalization of Neyman's Distributions. GEOFFREY BEALL.

It is known (Beall and Rescia, *Biometrics* 9: 354–386) that $\phi(t) = \exp. m_1 n! \sum_{s=1}^{\infty} m_2^s (e^{it}-1)^s/(n+s)!$ is the characteristic function of Neyman's contagious frequency distributions for $0 \leq n \to \infty$. Investigation now shows that n may properly assume negative values of any magnitude and thus produce an even more extensive family of frequency distributions. The distributions for $n=-1$ and $n=-2$ are intimately related to those where n is nonnegative in that, for example, the third moment steadily decreases from n infinite till $n=-2$. The distributions, however, change abruptly at $n=-3$; then the third moment becomes very great. As n decreases further the third moment again decreases and approaches the value obtaining for n very great. Presumably the Neyman distributions may now be expected to fit a great variety of data. It is known that with n small and positive they fit phenomena differing greatly from binomial situations. Current investigations suggest that much data that might be fitted by negative binomials can be as well fitted by the contagious distribution with $n \to \infty$. It may be hoped that the present extension will make them even more adequate. These further generalizations have suggested that the handling of this whole system of contagious distributions may be greatly facilitated by a transformation from the parameters m_1 and m_2 to $b = \frac{1}{2}(n+2)m_1/(n+1)$ and $c = 2m_2/(n+2)$.

The Time-interval Approach to the Problem of Contagion. GRACE E. BATES, *Mount Holyoke College.*

The particular approach to the problem of contagion discussed in this paper is an outgrowth of a suggested treatment outlined in the last section of the paper "Contributions to the Theory of Accident Proneness, Part II", by Jerzy Neyman and Grace E. Bates, U. of Cal. Press, 1952. Starting with a stochastic model in which the probability of an individual's incurring an accident in a given time interval depends not only on the length of this interval but also, possibly, on the number of previous accidents sustained—i.e. possible contagion—the random variables considered are the time intervals from the start of the period of observation to the occurrence of each accident for an individual incurring exactly n accidents in the interval, given that the individual had already sustained exactly m accidents at the start of the period of observation. In this paper, the distribution of these time intervals was considered only in the case of contagion of a very specialized type, termed linear contagion. Specifically, this restriction on the type of contagion provides, for example, that each additional increase in the number of accidents previously sustained (from 3 to 4, or from 10 to 11, for example) yields the same percentage increase or decrease in that individual's chances of avoiding accidents in the period of observation. Under this restrictive condition the distribution of time intervals is independent of the *number* of previous accidents sustained and the time-interval data needed for testing the hypothesis of no contagion does not require this information.

For this model, uniformly most powerful tests of the hypothesis of no contagion against either of the one-sided alternatives and a uniformly most powerful unbiassed test for the two-sided alternative case are obtained. The statistic used for these tests is the grand mean of all the time intervals for all the individuals. The exact power function can be given explicit form but is so cumbersome to apply that a method of approximating the power is outlined.

Contemporary Topics in Statistical Physics. G. W. PRESTON, *Philco Corporation.*

The history of the development of the classical theory of statistical mechanics is briefly traced. It is recalled that the behavior of macroscopic quantities of matter can be correctly described by the use of the Laws of Motion and the theory of probability. Not only the condition for thermodynamic equilibrium, but also the equations for the rate of processes, can be given by probabilistic statements. The fundamental statistical attributes of matter are the relative independence of the fundamental units of matter and their very frequent, though proportionately brief and violent interactions. The fact that the fundamental units of matter are nearly independent greatly simplifies the distribution problems, whereas their frequent interactions guarantee that the system will assume a very large number of microscopic configurations during the course of any thermodynamic measurement. The fundamental results of the recent theory of fluctuations are shown to suggest a possible approach to a statistical mechanical theory of non-equilibrium thermodynamics. Finally, the necessity of including in the expression for the physical entropy of a system the entropy of information obtained about the system by experimentation is shown to imply an equivalence between the second law of thermodynamics and certain basic theorems in the theory of statistical estimation.

Statistical Problems in Physics. MARTIN J. BERGER.

This paper reviews certain aspects of the relationship between physics and statistics. Some reasons are brought out why physicists in the past have paid comparatively little attention to problems of statistical inference. The point is made that this neglect is unjustifiable and deplorable. Various branches of statistics are then examined from the point of view of their importance to physics. Finally, some typical problems of statistical inference are described that arise in the interpretation of physical experiments. It is shown that there is a large class of problems in physics to which the existing statistical methods can be readily applied with great advantage, while others of a more specialized nature would require further development of statistical theory along novel lines. The hope is expressed that physicists will begin to make greater use of existing theory, and that statisticians will look into new problems raised in physics to the mutual advantage of both sciences

The Statistician in a Research and Development Laboratory. B. B. DAY, *U. S. Naval Engineering Experiment Station, Annapolis, Maryland.*

Using a particular research and development laboratory as a case history, the steps followed to put a statistician to work therein are outlined in detail. Three major points are developed: (a) the selling job required at the different levels with illustrative material presented; (b) the organizational set-up in the Laboratory for most effective work; and (c) the statistician on the job—qualifications, working relations, and responsibilities. Some details of the internal operation of a Statistical Office are indicated. The paper concludes with a discussion of what the future holds for the statistician in the Laboratory.

Survey of the Theory of Finite Sampling. JOSEPH F. DALY, *U. S. Bureau of the Census.*

Not many years ago the design of sample surveys depended heavily for its success on the reputations of its practitioners, who strove mightily to "validate" their samples with the aid of collateral information. Today, notwithstanding some popular impressions to the contrary, finite population sampling is one of the most respectable and thoroughly practical applications of mathematical statistics. The present quite satisfactory state of the theory of sampling finite populations is based on two main characteristics of the problems which the theory is designed to solve. In the first place, it is possible in many of these problems to devise methods of sample selection and estimation such that the resulting sample estimates can be expected to obey the laws of probability. The amount of variability in the estimates arising from the fact that they are based on a sample rather than on a complete count can therefore be measured by the same techniques and with the same precision that one can measure the variability in successive drawings from a table of randon numbers. In the second place, it is frequently possible to devise formulas which will serve as good approximations to the way in which the costs of projected sample surveys will vary with the size of sample and with the manner of selection of sampling units. This makes it possible to define objectively the notion of an "efficient" or an "optimum" survey design, namely one which minimizes the variability subject to fixed total cost.

Important work has been done in the way of devising techniques of sample selection which are more efficient than simple random sampling (e.g. selection with unequal probabilities, controlled random selection across strata, etc.) and on developing estimation formulas which make maximum use of available collateral information (e.g. regression estimates on one or more variables). One unsatisfactory aspect of the present theory is that it refuses to evaluate sample designs which are not subject to the laws of probability. Recent work on statistical decision theory based on the theory of games promises to shed some light on the properties of probability sampling methods in relation to larger classes of sample designs, including the selection of samples based on expert judgment. It is now known, for example, that under certain conditions simple random sampling represents a minimax strategy. Further developments along this line can be expected in the next few years.

Some Finite Sampling Concepts in Experimental Statistics. JEROME CORNFIELD, *National Institutes of Health.*

It is necessary to know the expectation of each of the mean squares in an analysis of variance in order (a) to choose a proper error term (b) to estimate components of variance. These expectations depend on what was sampled. Usually this dependence is expressed by denoting each observation as a linear compound of fixed and random variables, making certain assumptions about these components, and deriving the expectations. Different assumptions lead to different expectations, however, and in any analysis more complex than the one-way classification it is not always possible to choose unequivocally among the different possible assumptions.

A general way out of this difficulty is provided by finite sampling concepts. This is illustrated in the two-way classification, for which one can assume a population of elements classified into R rows and C columns with N elements in each of the RC cells. A sample of r rows, c columns and n elements within the rc cells in the sample is taken. If we assume the sampling is such that each row has the

same probability of selection (r/R), that each column has probability of selection c/C, and each element n/N, that the probabilities of sampling any two rows, columns, or elements within cells are respectively $r(r-1)/R(R-1)$, $c(c-1)/C(C-1)$ and $n(n-1)/N(N-1)$, and that rows, columns and elements within cells are sampled independently, the expectations can be derived without further (and more disputable) assumptions by elementary (but laborious) algebra to be

Sample mean square for	Expectation
Rows	$\sigma_e^2 \dfrac{N-n}{N} + n\dfrac{C-c}{C}\sigma_I^2 + nc\sigma_R^2$
Columns	$\sigma_e^2 \dfrac{N-n}{N} + n\dfrac{R-r}{R}\sigma_I^2 + nr\sigma_c^2$
Interaction	$\sigma_e^2 \dfrac{N-n}{N} + n\sigma_I^2$
Error	σ_e^2

where σ_e^2 is the within cell mean square for the population of RCN elements, $N\sigma_I^2$ the interaction mean square for this population, $NC\sigma_R^2$ the row mean square and $NR\sigma_c^2$ the column mean square. The expectations for Eisenhart's model I, model II, and mixed model follow as special cases.

STATISTICAL ABSTRACTS

All communications concerning this section should be addressed to the Abstracts Editor, Professor George E. Nicholson, Jr., Chairman of the Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Anis, A. A., and Lloyd, E. H., "On the range of partial sums of a finite number of independent normal variates," *Biometrika*, 40 (1953), 35–42.

Given a random sample of $n_{\text{s}}\{X_1\}$ from a normal population. The partial sums, $S_1, S_2, \cdots$, are formed, where $r=1, 2, \cdots, n$, $S_r=\sum_{i=1}^{r} X_i$. The average value of the range of the n partial sums is shown to be $\sqrt{2/\pi}\sum_1^{n-1} r^{-1/2}$. R. L. ANDERSON, *North Carolina State College*.

Anscombe, F. J., "Sequential estimation," *Journal of the Royal Statistical Society*, 15 (1953), 1.

The author reviews the literature on sequential estimation, presents some approximate methods for solving a number of particular problems and discusses the possible usefulness of sequential estimation procedures.

A sequential estimation procedure is a procedure in which the number of observations is not fixed in advance, but depends, according to some definite rule, upon the observations themselves. An example of such an estimation procedure, due to Haldane, is the estimation of a binomial proportion p when it is desired that the standard error of such an estimate be roughly proportional to p. The rule is to take observations until the number of successes X is equal to some prearranged number $X \geqq 2$. Then an unbiased estimate of p is $p = (X-1)/(N-1)$ where N is the total number of observations. An unbiased estimate of σ_p^2 is $p(1-p)/(N-2)$. Another example is that of estimating the mean of a normal population with unknown variance by a confidence interval of width d and confidence coefficient $1-\alpha$. Stein has given a double sampling procedure in which a first sample of fixed size n_0 is taken and then a further sample of $N-n_0$ observations is taken where N depends on the observations in the first sample. Reference is made to work of Girshick *et al.* for obtaining an estimate of the unknown proportion p at the end of a sequential test on a binomial population. The author asserts that estimation formulas valid for fixed sample sizes are asymptotically valid for sequential sampling if the sample size is large. He develops a sequential estimation procedure for estimating the mean of a normal population with unknown variance by a confidence interval of width d and confidence coefficient $1-\alpha$. If t_α is the normal deviate which has probability $\alpha/2$ of being exceeded, the rule is to stop taking observations when s^2 the unbiased estimate of the variance is first less than or equal to $l^2n^2/4t_\alpha^2$. Modifications of this rule are discussed which reduce the error of the estimate. It is shown that the expected sample size corresponding to this procedure is approximately $4\sigma^2 t_\alpha^2/l^2+1+t_{\alpha/2}^2$. Similar results are obtained for estimating the difference between two means, and estimating the birth and death rates in a simple birth-death process. Finally it is conjectured that perhaps the main practical benefit to be derived from studying the properties of sequential stopping rules is to assess the degree to which the usual fixed sample size estimation formulas are affected. GEORGE E. NICHOLSON, JR., *University of North Carolina*.

Altman, Irving B., "Relationship between sample size and AOQL for attribute single sampling plans," *Industrial Quality Control* (January 1954), 29–30.

A chart showing the AOQL values for a large variety of single sampling plans is presented as a function of sample size and acceptance number. GERALD J. LIEBERMAN, *Stanford University*.

Bailey, Norman T. J., "The total size of a general stochastic epidemic," *Biometrika*, 40 (1953), 177–85.

The author presents a stochastic treatment of the problem of estimating the number (w) who will become infected if one infected person is introduced into a group of n susceptible individuals. Diagrams are presented of the probability of w being infected for $n=10$, 20 and 40 and $\rho=n/4$, $n/2$ and n, where ρ is the ratio between the rate of removing infected people from the group and the infection rate. The average w is also computed. The results are extended to the distribution of w for house-

holds, when n is small; formulas are given for $n = 1(1)5$. The author summarizes his results as follows: "The model used here is not adequate for diseases with short infection periods, such as measles, but its adequacy for other infections requires testing." R. L. ANDERSON, *North Carolina State College.*

Bartlett, M. S., "Approximate confidence intervals," *Biometrika*, 40 (1953), 12–19.

The likelihood derivative, $T = \partial L/\partial\theta$, is used to derive approximate confidence intervals for various θ, assuming T is normal. An approximate correction for skewness (κ_1) of T is introduced, using $T_\lambda = T + \lambda(T^2 - I)$, where I is $\sigma^2(T)$. λ is approximated as $-\kappa_3/6I^2$. Also if $\kappa_3 = 0$, a method is given to adjust for kurtosis. Three examples are presented. R. L. ANDERSON, *North Carolina State College.*

Cohen, A. C., "Estimating parameters in truncated Pearson frequency distributions without resort to higher moments," *Biometrika*, 40 (1953), 50–7.

Using the method of moments, estimating equations are obtained which involve only the first four sample moments in contrast to the first six previously employed. The general case of doubly truncated four parameter distributions and special cases of doubly and singly truncated samples are worked out or their solutions indicated. It is pointed out that these equations can either be solved directly or by an iterative process. An example is worked out to illustrate the method and afford an evaluation of the technique. D. C. HURST, *North Carolina State College.*

Cox, D. R., and W. L. Smith, "The superposition of several strictly periodic sequences of events," *Biometrika*, 40 (1953), 1–11.

N sources, each producing events at regular intervals but with unequal periods, θ_i, $i = 1, 2, \cdots, N$, are pooled. This produces a sequence of unequal intervals. The frequency distribution for these intervals is derived. There is a point frequency, Q, for the largest of these intervals, which is the same as the smallest of the θ_i. A measure of the interval variability, displayed in a variance-time curve, can be used to distinguish the given sequence from a random sequence. Approximate procedures are given for large N.

Finally given a pooled sequence, a method is available to estimate the number of sources, N.

Three examples are given. This reviewer could not check the results for Q. R. L. ANDERSON, *North Carolina State College.*

Craig, C. C., "Some remarks concerning the Lot Plot Plan," *Industrial Quality Control* (September 1953), 41.

The author presents a non-technical evaluation of Dorian Shanin's Lot Plot Plan. He discusses three main points:

1. The misuse of lot plot for process control.
2. The severity of the operating characteristic curve of the Lot Plot.
3. The dependence on histograms. GERALD J. LIEBERMAN, *Stanford University.*

Craig, C. C., "Note on the use of fixed number of defectives and variable sampling sizes in sampling by attributes," *Industrial Quality Control* (May 1953), 83–86.

Under the assumption that samples are taken with replacements from small lots or that lots are large enough so samples taken without replacement will behave effectively as if they had been taken with replacements (infinite lots), the author examines the problem of attribute sampling where the random variable is the number of items sampled, n, before a fixed number, m, of defectives are found. A control chart based on the random variables n/m or n is presented. GERALD J. LIEBERMAN, *Stanford University.*

Craig, C. C., "On the utilization of marked specimens in estimating populations of flying insects," *Biometrika*, 40 (1953), 170–76.

An observer catches butterflies, marks them and then releases them. This is repeated s times in a given area and the numbers of butterflies caught one, two, $\cdots$ times are recorded. Five methods are used to estimate the total number (n) of butterflies in the area, three assuming a truncated Poisson distribution and two a special distribution due to Stevens. Approximate variances are obtained for each estimate.

Two worked out examples are included, plus the results of 14 other catches. The author uses χ^2 to test the agreement between actual and expected frequencies and finds 15 significance probabilities $>50\%$ and the 16th $>30\%$. R. L. ANDERSON, *North Carolina State College.*

Evans, D. A., "Experimental evidence concerning contagious distributions in ecology," *Biometrika*, 40 (1953), 186–211.

The frequencies for a variety of counts made on English and American plant species, insects and larvae and moth eggs are

fitted to three different theoretical contagious distributions: negative binomial, Polya-Aeppli and Neyman Type A. The mathematical formulae for fitting are included. For each distribution, two methods of computing are compared: (i) use of sample mean and variance, (ii) use of sample mean and proportion of zeros. A method of selecting between (i) and (ii) is presented. The Neyman Type A seemed to be best for the English plants, the negative binomial for the insects and moth eggs, and none of the three seemed to be satisfactory for the American plants.

Agreement between actual and expected distributions is tested by two statistics originally advanced by Anscombe. The usual χ^2-test for frequency data is also used, but this reviewer notes a tendency to pool too many classes on the tails. The Anscombe statistics seem to be more sensitive to aberrations from theory than is the all-purpose χ^2-test. R. L. ANDERSON, *North Carolina State College.*

Fisher, Walter D., "On a pooling problem from the statistical decision viewpoint," *Econometrica*, 21 (1953), 567–85.

The problem posed is that of finding a logical basis for grouping a set of K random variables into a set of G subgroups ($G \leqq K$). A procedure is outlined which places this problem in the framework of statistical decision theory. With a proper specification of the random variable and with the explicit recognition of additional assumptions and conditions, the logical subgrouping is determined from the observed sample data.

The following formulation of the general problem is offered. A sample ($x_1, \cdots, x_k$) consists of one observation drawn independently at random from each of K disjoint "cells." The random variable x_i has expectation θ_i, variance σ_i^2, and a measure of importance π_i. The θ_i are unknown, the σ_i^2 and π_i are known positive numbers. The problem is to arrange the set of K cells into G mutually exclusive subgroups according to some partition P, and to choose a decision vector $t_P = (t_{1P}, \cdots, t_{KP})$, where t_{iP} denotes a decision number or "estimated characteristic" of that group to which the cell i is assigned by partition P (t_{iP} for cells i assigned to the same group will be identical). Conditions may be imposed limiting the number of subgroups G and also the partitioning procedure. There remains a number of theoretically admissible alternative partitions and a number of possible decision vectors for each partition. A choice of the optimum decision vector t^* involves the multiple choice of an optimum partition P^* and an optimum set of decision numbers t_{iP}^*. Following decision theory a loss function $W(\theta, t_P)$ is defined which measures the cost incurred by making the decision t_P when the true parameter point is θ. The vector t^* is determined as the set of decision numbers which minimizes the expected value of W over the domain of possible samples.

The procedure is formulated in specific terms with reference to the following hypothetical problem: A public price control agency is assumed to have the task of setting selling prices for a given commodity. The commodity is sold in K independent markets, each with a linear demand function of the form $p_i = 2\theta_i - (1/\pi_i)q_i$, where p_i is price, q_i is quantity sold, and π_i is a positive number, ($i = 1, \cdots, K$). It is assumed that the price control agency in setting prices is guided by two conflicting policies: (1) to be "fair" to sellers, and (2) to attain low administrative costs. Administrative costs are assumed known and are assumed to increase with the number of different prices set by the agency. A measure of "unfairness" to sellers is assumed in the present problem to be represented by the departure of actual realized returns under adminis ered prices from maximum realizable returns under free market prices. Under the conditions assumed the expression for the loss function becomes $W(\theta, t) = c_G + \sum_{i=1}^{K} \pi_i (t_{iP} - \theta_i)^2$, where the first term on the right represents total administrative costs incurred if G different prices are set and the second term represents aggregate returns sacrificed by sellers as a result of the administered prices t_{iP}.

Apart from the price agency illustration but assuming a loss function of the same specific form, stochastic properties are ascribed to the set θ and a risk function (expected value of the loss) defined. A limiting Bayes solution to the decision problem is derived (and presented in an appendix) under the assumption of a sequence of a priori distributions of θ having constant density and increasing range. It develops that the optimum decision procedure t^* is the following: For a sample x (in the price agency problem x_i is estimated equilibrium price in market i) select for each value of G (number of different administered prices) the admissible partition P_G^* of the K cells (independent markets) which minimizes the quantity $S_P = \sum_{i=1}^{K} \pi_i (x_i - \bar{x}_{iP})^2$, where $\bar{x}_{iP}$ is a weighted sample group mean. The weighted means of the sample observations within groups formed by the optimum partition for a specified G give the optimum set t_G^*. Denote the above sum of squares

corresponding to P^* for a specified G by S_G^*. Then the optimum set t^* is given by G which minimizes $(c_G+S_G^*)$.

This solution is applied to sample data on retail prices of fresh tomatoes in nineteen independent retail markets in California. In this numerical example restrictions are placed on the partitioning procedure which limit the number of admissible partitions for each G greater than 1, and G itself, limiting the maximum number of different administered prices to 4. In addition, a schedule of administrative costs c_G is assumed for $G=1, \cdots, 4$, and π_i is assumed uniform for all markets. The resulting optimum set of administered prices is presented. It turns out that the optimum number of different prices in this case is 4, the maximum allowable.

In a concluding part the author offers some comments with respect to limitations and possible extensions of the proposed pooling procedure and discusses briefly the relation of the pooling problem to certain other subjects such as the problem of aggregation discriminant analysis, classification problems, and factor analysis. IVAN M. LEE, *University of California.*

Fox, Karl A., "A spatial equilibrium model of the livestock-feed economy in the United States," *Econometrica*, 21 (1953), 547–66.

In this paper the author obtains numerical solutions for a ten-region spatial equilibrium model for feed grain of the type formulated by P. A. Samuelson in "Spatial Price Equilibrium and Linear Programming," *The American Economic Review*, June, 1952. The problem is not formulated explicitly in a *statistical estimation* framework, although actual data and approximate relations derived from actual data are employed in the computations. The paper serves primarily to illustrate unique numerical solutions to the spatial equilibrium model for a particular simplified formation of the problem and under alternative assumptions regarding the magnitudes of certain variables entering the model.

Quantitative results showing regional equilibrium prices, net trade, and regional patterns of trade for conditions with respect to freight rates, feed supplies, livestock numbers, and livestock prices approximating those for 1949–50 are shown. Solutions are also shown for two alternative sets of conditions: (1) Conditions like 1949–50 except that freight rates are assumed uniformly 50 per cent higher and (2) conditions like 1949–50 except that feed supplies are assumed lower in the Corn Belt and Northern Plains regions by 30 per cent and 75 per cent, respectively. An "application to forecasting" is also illustrated, which differs from the problem solved for 1949–50 conditions only in that July, 1947 forecasts and December, 1947 estimates of feed grain production are used in arriving at regional feed grain supplies. IVAN M. LEE, *University of California.*

Gulde, Harold J., "Acceptance sampling by variables using the range," *Industrial Quality Control* (November 1953), 18–25.

Formulas and charts are presented for constructing variables sampling plans using the range and average range. The results are based upon the Normal approximations. GERALD J. LIEBERMAN, *Stanford University.*

Horsnell, G., "The effect of unequal group variances on the F-test for the homogeneity of group means," *Biometrika*, 40 (1953), 128–36.

The author extends the 1951 investigations of David and Johnson (*Biometrika*, 38: 43), using four groups in a single classification analysis of variance. Most of the computations are based on three equal within-group variances with a fourth variance three times as large. Assuming equal group means, the actual significance probabilities corresponding to nominal 5% and 1% significance levels of F are obtained for equal and various combinations of unequal numbers of observations per group.

The power of the usual F-test to detect one divergent group mean (out of the four) is approximated both when the variable group has and has not the divergent mean.

The author summarizes his results as follows: If there is no very clear information as to heterogeneity, use equal group frequencies. If it is known that one group is more variable, make sure that this group does not have less observations than the others; if possible, take a few extras in this group.

An alternative to the usual F-test is also mentioned. R. L. ANDERSON, *North Carolina State College.*

Kamat, A. R., "On the mean successive difference and its ratio to the root mean square," *Biometrika*, 40 (1953), 116–27.

Given a sequence of n normal variates $\{X_i\}$ with means $\{\mu_i\}$ and common variance σ^2. The mean successive difference is

$$d_s = \sum_{i=1}^{n-1} \frac{|x_i - x_{i+1}|}{n-1}.$$

The first four moments of d/σ are derived when all $\mu_i=\mu$. The standard deviation, β_1 and β_2 values of d/σ are tabulated for $n=3$ (1) 10(5) 30, 40, 50; a type I Pearson curve was used to obtain approximate upper and lower $\frac{1}{2}$, 1, $2\frac{1}{2}$ and 5% percentage points. Exact results are available for $n=3$. Empirical sampling was used to check some results.

The usefulness of d over the standard deviation, s, is shown when the μ_i shift.

The distribution of $W=\bar{d}/s$ is considered when all $\mu_i=\mu$; its mean, standard deviation, β_1 and β_2 values are given for $n=5(5)30$, 40, 50. Approximate percentage points are also presented. R. L. Anderson, *North Carolina State College.*

Kamat, A. R., "Incomplete and absolute moments of the multivariate normal distribution with some applications," *Biometrika*, 40 (1953), 20–34.

Given n normal variates, $\{x_i\}$, with zero means and given variances and covariances. Absolute moments are moments of $\{|x_i|\}$. Exact results are given for $n=2$, 3, and power-series for derivations when $n>3$. These results are applied to the distribution of $T=\sum|x_i|$, using Pearson-curve approximations. R. L. Anderson, *North Carolina State College.*

King, E. P., "Estimating the standard deviation of a normal population," *Industrial Quality Control* (September 1953), 30–33.

The efficiency of various estimates of the standard deviation of a normal population is presented when nk observations are grouped into k subgroups of n observations. Gerald J. Lieberman, *Stanford University.*

Latscha, R., "Tests of significance in a 2×2 contingency table: Extension of Finney's table," *Biometrika*, 40 (1953), 74–86.

Finney's 1948 table (*Biometrika*, 35: 148) for making exact tests of independence in a 2×2 contingency table with fixed border totals is extended through $A=20$. This table is of the form

	Number of Successes	Number of Failures	Total
Series I	a	$A-a$	A
Series II	b	$B-b$	B
Total	$r=a+b$	$A+B-r$	$A+B$

The experimenter defines a success and Series I such that $A\geq B$ and $a/A\geq b/B$. For a given A, all combinations of B and a are included which will give a significant result, where the significance probabilities are $\alpha=.05$, .025, .01 and .005 (one-tailed). For each combination of A, B and a, the author gives the value of b ($<a$) which is just significant at the α-level, b_α, and the exact probability that $b\leq b_\alpha$, for a given A, B and r. R. L. Anderson, *North Carolina State College.*

Lindley, D. V., "Estimation of a functional relationship," *Biometrika*, 40 (1953), 47–49.

It is desired to estimate α and β in the linear model $V=\alpha+\beta U$, where the only observations on U and V are subject to error and all distributions are normal. U and V are allowed to be either mathematical variables or random variables. If the observations on one of the variables are at predetermined fixed points, the so-called "controlled" observations, it is demonstrated that the estimating equations for α and β can be obtained in the usual way. This result also applies if these "controlled" observations are allowed to be random. If there is no "control" on either set of observations, the estimation is deemed impossible without recourse to supplementary information. D. C. Hurst, *North Carolina State College.*

Moore, P. G., "A sequential test for randomness," *Biometrika*, 40 (1953), 111–15.

Given a sequence of observations, where each observation falls into one of two alternative categories. A sequential procedure is presented for testing the hypothesis, H_0, that these observations occur in random order, against the alternative H_1, that there is dependence of the kind found in a simple Markoff chain. The procedure is illustrated with annual rainfall data. C. E. Gates, *North Carolina State College.*

Scheffé, Henry, "A method for judging all contrasts in the analysis of variance," *Biometrika*, 40 (1953), 87–104.

The author presents and proves the validity of a method for making further inferences about the contrasts among a set of true means or main effects $\mu_1, \mu_2, \cdots, \mu_k$, following the rejection of the hypothesis $H:\mu_1=\mu_2=\cdots=\mu_k$ by the conventional F-test with $k-1$ and ν degrees of freedom in the analysis of variance. The method is based on a probability statement concerning the infinite totality of contrasts of the type

$$\theta = \sum_{i=1}^{k} c_i \mu_i$$

where the c_i are any known constants, satisfying the condition

$$\sum_{i=1}^{k} c_i = 0.$$

If the assumptions usual in the analysis of variance hold and $\hat{\theta}$ and $\widehat{\sigma_{\hat{\theta}}^2}$ denote the estimates of θ and the variance of $\hat{\theta}$, then it is proved the probability is $1-\alpha$ that the values θ of *all* possible contrasts simultaneously satisfy

$$\hat{\theta} - S\widehat{\sigma_{\hat{\theta}}} \leq \theta \leq \hat{\theta} + S\widehat{\sigma_{\hat{\theta}}} \qquad (1)$$

where S^2 is $k-1$ times the upper α point of the F-distribution with $k-1$ and ν degrees of freedom. The result holds for any values of the unknown parameters. The result may be used for the interval estimation of any contrast of interest or to declare any estimated contrast "significantly different from zero" or not, according as the corresponding interval (1) excludes $\theta=0$ or not, contrasts suggested by the observed means $\hat{\mu}_i$ included.

A method has been prescribed by Tukey for the special case where all the $\hat{\mu}_i$ have the same variance and all pairs $\hat{\mu}_i$, $\hat{\mu}_j$ $(i \neq j)$ have the same covariance, and where there is interest in only a subset of the totality of contrasts, namely the $(1/2)k(k-1)$ differences $\mu_i - \mu_j$. This latter method is shown to have greater sensitivity in this situation than the method based on (1) in the sense that the confidence intervals are shorter.

The author examines the operating characteristic of the method by determining (i) the probability that all contrasts whose true values are zero would be declared "not significantly different from zero," and (ii) the probability that all normalized contrasts whose true values are greater in absolute value than some specified bound will be declared "significantly different from zero." D. G. Horvitz, *North Carolina State College.*

Stevens, W. L., "Tables of the angular transformation," *Biometrika*, 40 (1953), 70–73.

The angular transformation is defined as

$$\theta = 50 - \sqrt{1000} \text{ arc sin } (1 - 2p),$$

where p is the estimated proportion for a binomial population. Values of θ are given to 3 decimal places for

$$p = 0(.0001).02(.001).5,$$

and for selected proper fractions. R. L. Anderson, *North Carolina State College.*

Stuart, A., "The estimation and comparison of strengths of association in contingency tables," *Biometrika*, 40 (1953), 105–10.

A modification of Kendall's rank correlation coefficient, t_c, is proposed for measuring strength of association between two characteristics in an $r \times s$ contingency table. Conservatively approximate confidence limits are set for the population association. An approximate test is provided for the difference between the coefficients calculated by this method for two $r \times s$ contingency tables. The coefficient used is

$$t_c = \frac{n-1}{n} \frac{m}{m-1} t_s$$

where t_s designates Kendall's coefficient of rank correlation. S. Weiner, *North Carolina State College.*

Tanner, J. C., "A problem of interference between two queues," *Biometrika*, 40 (1953), 58–69.

Imagine a bridge of certain length AB only wide enough to admit a single lane of traffic, with traffic arriving at both ends. Vehicles V_1 arrive at A at random and at an average rate q_1 per unit time. A particular V_1 takes time α_1 to cross AB, but may enter AB only provided no V_2 vehicle is in AB, and if no other V_1 vehicle has entered AB during the previous time β_1. Similar sets of conditions, α_2 and β_2 affect the V_2 vehicles arriving at B. The mean delay time $\bar{w}_1$ and $\bar{w}_2$ for $\alpha_1 > \beta_1$ and $\alpha_2 > \beta_2$ is obtained as a function of t_1 and t_2, the time for a block of V_1 or V_2 vehicles to cross, and r_1 and r_2, the number of vehicles waiting to enter the bridge. Explicit solutions are made available for the special cases $\beta_2 = 0$ and $\beta_1 = \beta_2 = 0$. Further applications to the problems of delays in crossing or entering a heavily travelled artery are described, and delays due to intersecting lines of traffic discussed. Two short tables are included, giving the values of $\bar{w}_1$ for $\beta_1 = \beta_2 = 0$, $\alpha_1 = \alpha_2 = 1$ and for $\bar{w}_2$ when $\alpha_2 = \beta_2 = 0$, $\alpha_1 = 1$. J. S. Hunter, *North Carolina State College.*

Whitfield, J. W., "The distribution of total rank value for one particular object in m rankings of n objects," *The British Journal of Statistical Psychology* 6, part 1 (1953) 35–40.

The title clearly indicates a class of problems not appropriately handled by Kendall's treatment of the general m ranking problem. As an illustration one can consider a situation in experimental social psychology where one person in a group is instructed to play a predetermined role and each of

the other members of the group is asked to rank his fellow members (including the experimental person) in connection with certain characteristics. Under the assumption that for the first ranking each value $(1, 2, \cdots, n)$ has equal probability the frequency distribution for m rankings of n objects is constructed; the mean total rank value is $\frac{1}{2}m(n-1)$, its variance is $[m(n^2-1)]/(12)$, $\beta_2 = 3-(6)/(5m)-(12)/[5m(n^2-1)]$. Tables of exact probability values are given for m and n up to 8. For higher values of m or n $[|\text{Total Rank Value} - \frac{1}{2}m(n-1)| - (\frac{1}{2})]/\sqrt{[m(n^2-1)]/(12)}$ is approximately a normal deviate. For example, if eight judges rank seven objects, and the experimental object has a total rank value of 43; the approximation gives $P = .0318$ against the exact value of .03113. HERBERT SOLOMON, *Columbia University*

BOOK REVIEWS

Introduction to the Theory of Statistics. *Victor Goedicke.* New York: Harper and Brothers, 1953. Pp. xii, 286. $4.50.

BERNARD L. WELCH, *University of Leeds, England*

THE present text sets out the fundamentals of statistical method in a sound and readable manner and can be recommended to students who are starting upon the subject from the very beginning.

Since no use is made of the calculus the author is often forced to be purely descriptive and to dispense with much theoretical justification. This seems to be inescapable when only a small prior knowledge of mathematics is assumed. The reader is not even required beforehand to be familiar with the use of logarithms and with simple graphs. Sections are devoted to these topics at appropriate points as they are needed. I feel, myself, that writers of elementary books on statistics should not be expected to go so far as this in remedying deficiencies in the mathematical equipment of their readers. If a student has not already at least some facility with graphs and logarithms then he is, I believe, ill-advised to start to grapple with the theory of statistics at all.

The decision of the author to include sections on permutations and elementary probability, however, is not open to the same criticism and his introduction to these topics is admirable. In his treatment of frequency distributions, the expression of frequency as a proportion per t unit (where $t=(x-\bar{x})/\sigma$) leads on naturally to the standard normal curve as a graduation of frequency distributions arising in many different practical contexts. Some theoretical justification of the general use of the normal curve is attempted by including a numerical investigation of the limiting behavior of the binomial distribution. This section, however, makes more difficult reading than the rest of the book and could possibly be simplified in places.

The treatment of correlation is excellent, the emphasis rightly being placed on the square of the correlation coefficient rather than r, itself—the idea being to stress that the important factor is the fraction of the variance accounted for by a straight-line relation. Multiple correlation is dealt with again from the same angle and the examples used to illustrate it are well-chosen.

The introduction to the sampling variability of such simple statistics as a sum, a difference, and a mean value is clear and leads on to the general problem of testing statistical hypotheses where, however, some statements are made which are open to criticism. For instance in an inquiry where the effects of two drugs on the period of convalescence after illness are being compared, the author gives an example where the probability is 0.0018 that a difference between drug-means greater than the one observed could have arisen on the null hypothesis. He goes on to deduce that "the available evidence indicates that the probability that the drug does *not* affect the duration of convalescence is only 0.0018, while the probability that it does affect duration is 0.9982." Such a statement as this is certainly unwarranted and some-

thing more by way of explanation seems to be called for at this point. One must grant, of course, that it is somewhat difficult to explain in an elementary fashion what the interpretation of the figure 0.0018 should be. The devotion of more space to the clarification of this general problem of testing hypotheses would not be a waste, however, for as they stand, the sections on this topic suffer by reason of compression.

The concluding section of the book draws attention to certain common-sense viewpoints which the novice is apt to overlook.

The standard of printing is high throughout and altogether this is a useful addition to the texts already available on the elementary parts of statistical theory.

Statistical Inference. *Helen M. Walker* and *Joseph Lev.* New York: Henry Holt and Company, 1953. Pp. xi, 510. $6.25.

PALMER O. JOHNSON, *University of Minnesota*

THIS text is written primarily for non-mathematical students. The complete text is planned for a course two-semesters in length. The problems connected with writing text-books for the type of students specified are particularly difficult if the responsibilities of the author(s) are to be fulfilled. There is the problem of what mathematics can be assumed on the part of the reader. A second problem resides in the selection of content. Since texts of the kind discussed here are written for those who are to become practitioners or research workers, the most important statistical procedures must be described and the computational problems arising must be discussed. The methods must also be illustrated with representative examples from the field or fields of application, usually with the particular field with which the author is familiar. There is likewise the selection of problems to be solved by the student for testing his understanding of the theory and methods. The writer of texts for non-mathematical students is under particular obligation to describe rigorously and to indicate in his own illustrations the assumptions to be fulfilled if the procedures are to be validly applied. This is particularly important in a book whose readers lack the knowledge necessary to check the author's statements.

Unfortunately there are no properly designed experiments to test the efficacy of text-books on modern statistical methods. It is, therefore, difficult to make criticisms that are completely objective. Most statisticians have strong convictions not only on the relative importance of various statistical procedures but also on how mathematical (even for non-mathematical students) a text-book should be.

In the reviewer's judgment, the authors of this text have in general succeeded in writing a text that reaches a high level of attainment in meeting the criteria suggested above.

The mathematical requirements would be chiefly met by a good command

of arithmetical and algebraic processes. There is considerable use of probability symbolism, multiple subscripts, multiple summation, and specified limits of summation. In this connection, well stated explanations and practice exercises in the processes involved are provided. With respect to previous statistical training, the student with good ability and familiarity with symbolism can acquire the elementary calculation processes, such as calculation of the mean, standard deviation, regression, and correlation from the treatment in the text.

Skilful use is made of intuitive reasoning and of effective graphical devices for developing a functional understanding of statistical concepts and processes underlying statistical inference.

The nature of statistical inquiry is introduced in Chapter 1 by a simple experiment. Chapter 2 lays the basis for a more formal treatment of probability and its role in statistical inference. Fundamental concepts are introduced, such as variable, population, probability, random observation, probability distribution, independence, elementary laws of probability, parameter, and statistic. Particular application is made to the binomial distribution. Chapter 3 extends the concepts and methods of statistical inference to the types of problems encountered in the use of the binomial model. An unusually clear exposition is given of such fundamentals as sampling distributions, critical regions, levels of significance, point and interval estimation, tests of alternative hypotheses, power of a test, one- and two-tailed tests, and the two types of errors. Turning to Chapter 4 an extension is made of the previously developed concepts from the binomial to the multinomial populations comprised of several discrete classes. The content deals mainly with chi-square, its sampling distribution, properties, the chi-square curves and tables. A number of applications are made, emphasizing calculations and interpretation. No mention is made of the criticisms and limitations of the chi-square test of "goodness of fit," particularly that the power of the test to detect disagreement between hypotheses and observations is determined largely by the size of the samples.

In Chapter 5 and in subsequent chapters consideration is given to problems in inference involving continuous variables. Certain general concepts underlying distribution of a continuous variable are developed. Notions of the parent population, unbiased estimates, and a test of normality are included. Chapter 6, on sampling distributions, presents a number of empirical distributions of commonly used statistics. The frequency distributions of various statistics are plotted and the mathematical curves superimposed. Consideration is given to statistics following the chi-square, student's t, and the F-distributions. The property of independence of statistics is well portrayed by plotting the joint frequency distribution of the means and standard deviations of samples from a normal population. Chapter 7 presents methods for testing hypotheses and for estimating population means and differences between means of two populations. The test of alternative hypotheses is

especially well treated including the graphical interpretation of the power of a test. Limited consideration is given to the design of samples in surveys. Estimates for the mean and variances are given for stratified and cluster samples. Along similar lines Chapter 8 considers the problems of inference concerning variances and standard deviations of normal populations. Special consideration is given to the use of the tables of the F-distribution. Bartlett's and Hartley's tests for the equality of variances of several samples are applied to practical problems. There is also a test for differences between variances of correlated measures.*

Chapter 9 presents some of the simpler applications of the analysis of variance. A k-sample problem of testing the equality of means is illustrated in an experiment designed to determine if order of shooting three different ranges in archery had any effect. The underlying mathematical model is specified including assumptions. In this connection it would have been desirable to carry out the tests of normality and of equality of variances rather than merely to indicate that these assumptions had been satisfied. The test of the equality of means is carried out by estimating the unknown population variance from the variation among the three sample means and from the variation among the archers within groups, then by taking the ratio of the two estimates, the variance ratio, or F, as the test statistic. This approach makes it difficult for students to perceive why these sums of squares were selected. The approach through stating the mathematical model in algebraic form or in the form of a linear hypothesis seems to be more informative. It would entail the use of the method of least squares or of maximum likelihood to obtain the appropriate sums of squares. In this chapter there is an excellent explanation of the form of the F-distribution, including curves of the sampling distribution of F and $1/F$.

In Chapter 10 methods of statistical inference are applied to bivariate data. A mathematical model for linear regression and the sampling distributions of statistics in linear regression are presented as the bases for testing hypotheses and for interval estimation of parameters. The bivariate population model and the normal bivariate population model are treated. The distribution of the product moment correlation coefficient is described. Tests of hypothesis about ρ are carried out and confidence intervals for ρ set up. There is another chapter (Chapter 11) given to other measures of relationship. The chief contribution here is a test of significance for point-biserial-r and a brief summary of Tate's unpublished study comparing bi-serial and point-biserial-r.

Chapter 12 is devoted to the statistics of measurement. This chapter should serve to bring to the attention of the research worker the importance of errors of measurement particularly in tests of significance and in estimation. Most of the content of this chapter is of long standing and indicates how little workers in this field have been influenced by the developments in modern statistical inference. Most of the old algebraic formulations are

reported in this chapter. Perhaps an opportunity would have been afforded to report some beginnings that have been made and other help that could be had from ideas and developments in statistical inference. Examples are Wilk's L_{mvc} criterion for parallel tests, Votaw's test of compound symmetry, component mean square analyses in regression problems, the practical sampling problems of reliability and validity, and the uses that are beginning to be made of information theory in mental test theory and construction.

Chapter 13 gives a lucid explanation of the methods and interpretation of multiple regression and correlation. Illustrations are given of the Doolittle method and Fisher's modification of the Doolittle method in obtaining a multiple regression equation and tests of significance. There is no consideration in this or other chapters of other methods of multivariate analyses. Thus the discriminant function designed for obtaining the best system of weights of various independent variables for distinguishing among groups of individuals is not treated, nor are such methods discussed as Hotelling's T for testing the significance of the difference between means of multivariate normal populations, and methods of classification of an individual based on a number of measurements into one of several categories. Examples of the latter methods are the generalized distance function and the statistical decision function.

In Chapter 14, the analysis of variance is extended to the analysis of two or more variables of classification. This chapter bases its discussion around data derived from studies employing modern experimental designs. In following through the analysis of the designs presented, the student should gain a realistic conception of the role of statistics in the analysis of the data from modern designs. It would have been desirable to have given more explanation to the role of statistics in the planning stage of the design and greater emphasis to the functions of randomization and replication.

A few points of criticism will be made to an otherwise admirable chapter:

> On p. 349, μ in H_r and H_c appears to be the same value, which in general is not true unless the value is zero. On p. 354, the test of the hypothesis that the mean of the 4 diagonal cells is equal to the mean of the 12 non-diagonal cells by means of the t-test is not valid unless probability corrections are made. The latest proposals for the solution of this problem are the analyses suggested by Scheffé and by Tukey. There is no explanation in the text of the origin of expected mean squares in tables 14.6 and 14.12. Nor is any statement made tying these values up with Model II of the analysis of variance. On p. 363, recognition is given to the circumstance that when several measures are obtained from the same individual, the measures are likely to be correlated. It is not clearly indicated how, if at all, the effects of the correlated observations have been removed.

Chapter 15 gives special consideration to the use of the analysis of covariance in group comparisons on one variable when information is available on another or on several other variables which are correlated with it. Skilful demonstration is given by the analysis of data from a number of important research problems in this area.

In Chapter 16, certain special uses of percentiles are illustrated in analysis

leading to inferences. In Chapter 17 several useful transformations are discussed. The final chapter discusses and illustrates the use of a number of non-parametric methods in testing hypothesis and in setting up confidence intervals.

The book has an exceptionally large number of useful tables and charts (20 in all), besides a table of squares, square roots, and reciprocals, a table of four-place logarithms, and one of random numbers.

There is a five-page glossary of symbols. A detailed table of contents and a well-prepared index facilitate the location of items of interest. All but four of the chapters contain a list of references.

Eleven of the chapters contain exercises for the student. They are most frequently interspersed within the chapter. These exercises are well designed to test understandings of terms, phrases, principles, symbols, formulas, use of tables, etc. Only three chapters contain exercises dealing with real data. To the reviewer, it seems that in courses for practitioners abundant opportunity should be given to deal with real data—fictitious data do not often give experiences that will be met with in research. There are no such opportunities in Chapters 14, 15, 17, 18, for example, which deal with the types of situations most often encountered in practice or likely to be used by students in their own research. These types of situations are likely to be most motivating and to develop an appreciation of the dynamic qualities of statistics.

In summary, the reviewer believes that this text will make a valuable contribution by elevating the plane of instruction in applied statistics. While the fields of application are chiefly in education and psychology, the book could profitably be used as a basic text in other fields. It will particularly appeal to instructors in the social science fields because so many of the current texts are written for students in biology and in agriculture.

Statistical Methods in Experimentation: An Introduction. *Oliver L. Lacey.* New York: The Macmillan Company, 1953. Pp. xi, 249. $4.50.

OSCAR KEMPTHORNE, *Iowa State College*

THIS text is designed for use in a one-semester course in general statistics as applied to experimentation, and assumes one semester of college algebra as a background. The subject matter of the book is indicated by the chapter headings: The aim and problems of statistics in experimentation; Experimental design; Interpretation; Probability (I)—the probability of discrete events; Probability (II)—probability in a continuum; Three chapters on the normal distribution; Tests of significance of means and differences between means; Enumeration data; Correlation; Regression; Fiducial limits; Experimental design; Appendix of tables.

The plan of each chapter is to give first a general discussion of the matter under consideration, then examples, which are in the form of questions, and

answers to these examples, and finally problems for the student. The general form of the book, the level of discussion of topics, and the question-and-answer technique strike this reviewer as being very good and very appropriate to a first course in statistics for students with no mathematical background. The format and general style are very clear and should not be a source of confusion or annoyance to the intended readers.

A number of criticisms can, however, be made.

With respect to the design of simple comparative experiments, the discussion is somewhat naive. The author states that there are four general methods of controlling variation due to extraneous factors (1) elimination (2) equalization, (3) balancing, (4) randomization. The first two are never completely realizable, and the third is only realizable with special assumptions which may be very arbitrary. The basic fact is that the first three methods are valuable adjuncts to randomization and it is rare that randomization can be dispensed with. In any case if the physical act of randomization is not done the validity of the statistical inference is questionable as Fisher originally stated. The combination of system and randomness in a design is little discussed, and this is perhaps the most important feature experimentally.

The discussion of the size of experiment is made generally in terms of "minimal adequacy," although the word efficiency is used occasionally. It would seem desirable pedagogically to conform to the usual use of the word "efficiency," and to retain the notion of power instead of introducing a new notion.

It would be desirable that all statistical texts make the distinction between regression relationships and functional relationships. Originally regression was a relationship in a multivariate population, but now is widely used for the relationship of y to x, where x is not a random variable. When one uses regression analysis on the growth curve of plants, one is really estimating a functional relation of growth (y) to age (x) and in this case there is no point in considering an underlying population of x values from which one has a sample. In a true multivariate situation one may be interested in correlation or regression, but in the functional relationship situation the correlation has no real significance. All the problems in the chapter on regression are the functional type. Uses of regression in multivariate populations could be discussed more fully, and examples given.

The author acknowledges the stimulation of R. A. Fisher's *The Design of Experiments*, but certain passages indicate that there is some confusion in the author's mind on some basic questions. For example he gives as a sample question, "Does home economics training result in a better chance for marriage than arts college training?" And then gives the answer "This may be attacked experimentally by keeping comparative records for a number of years of home economics and arts college graduates. Obviously care would be needed to make groups truly comparable in respects other than academic training." The reviewer disagrees entirely with this answer; the answer that

is given is not an answer to the question posed. The question posed can be answered strictly by giving girls home economics training or arts college training at random, without regard to their inclinations or anything else and then observing their marital success. The answer which would be obtained by the stated procedure is the answer to the question "Do home economics graduates have a better chance for marriage than liberal arts graduates?" The lack of appreciation of the difference between the question posed and the question answered appears to be rather widespread in the social sciences. Also there is a question about the answer: what is meant by "truly comparable?" The adjective "truly" appears to have taken on some mystic significance with some scientists and the use is better avoided. In addition the reviewer does not like the use of the word "experimentally" in the answer, in that one is merely observing an uncontrolled situation. Nor does the reviewer favor the use of "identical" in statements such as "we should present them (*two drinks*), therefore, in identical glasses" p. 3 (cf. the use of "equal" on page 18).

Writing really good elementary texts in statistics is a very difficult job. The present book is a good attempt, and would be more than good if the defects mentioned above were corrected.

Methods of Statistical Analysis in Economics and Business. *Edward E. Lewis.* Boston: Houghton Mifflin Company, 1953. Pp. viii, 686. $5.50.

Z. SZATROWSKI, *University of Buffalo*

IN THE opinion of the reviewer, Professor Lewis has written a definitely good book on statistical methods in business and economics. He has done this by adding to the conventional material some of the fundamental concepts of modern statistical analysis. His discriminating selection of subject material and careful allocation of space according to the importance of basic topics enables him to cover more and get it across. The exposition (text, illustrations, and problems) is excellent.

The organization of the book is reasonably conventional and departures from the pattern of the past are improvements. There are 686 pages, five parts and an appendix. Part I (pages 1–96), the introduction, discusses the purpose and problems of statistical methods. Also this section includes an explanation of graphs, tables, and "Statistical Numbers and the Problem of Accuracy." Part II (pages 97–186) presents descriptive statistics for frequency distributions, i.e., measures of central tendency including the geometric mean, measures of dispersion, skewness, and kurtosis. The application of these descriptive statistics is explained clearly. The illustrations involving comparisons are very effective. In connection with this section, the reviewer would like to call attention to an inaccuracy in the specification of class limits and mid-points. The author's presentation results in a bias. To illustrate, on page 67, Table 6, for weekly earnings recorded to two decimals,

the first class is given as 22.00–23.99 with mid-point 23.00. More accurately it should be 21.995–23.995 with mid-point 22.995.

Part III (pages 187–298) is entitled "Statistical Inference." Here are contained various tests of significance, for small as well as large samples, including the application of the F and χ^2 Test. This section concludes with a discussion of quality control methods. In the reviewer's opinion, this is an outstanding presentation. The problem of inference is given proper emphasis. The various applications, which in so many texts are scattered throughout the book, now appear in one section. The author succeeds in covering a great deal in this part because he has a logical unified organization of the material and because he develops each well-chosen illustration of inference just far enough to present the idea clearly.

Part IV (pages 299–370) very adequately discusses the subject of index numbers, their construction, the problems, and current applications. Part V (pages 371–486) deals with time series analysis, trends, cycles, and seasonals, including the changing seasonal. This section is not involved but thorough. In connection with trend determination, the author includes a brief discussion of the use of transformations, so widely applicable in non-linear trend problems. Part VI (pages 487–649) is a clear presentation of correlation, multiple as well as simple, including tests of significance.

The Appendix in this book contains a list of references, formulas, and tables of square roots and logarithms. Tables of the Normal, t, F, and χ^2 distributions appear in the text where applications are discussed.

The book contains an adequate number of problems which appear throughout the chapters to illustrate specific topics which the author has just discussed. In general, the problems are well chosen and involve the minimum of calculation. The format of the book is in keeping with the high quality of the contents.

An instructor using this book can expect his students to get much out of the text. Also he can expect many of his students to become interested in statistics. This book should enjoy success as a text. In addition, since it is easy to understand and relatively complete and up-to-date in its coverage, it should serve as a very useful reference book in business and economic statistics.

Sampling Inspection by Variables. *A. H. Bowker* and *H. P. Goode*. New York: McGraw-Hill, 1952. Pp. xi, 216. $5.00.

H. C. HAMAKER, *Philips Research Laboratories, Eindhoven, Holland*

THE first thing one notices about a new book usually is its title, and it is therefore important that this title should convey a concise and correct impression of the content of the book it covers. The title of the book under review does not quite satisfy this criterion and I am afraid that many statisticians who ordered the book because of its attractive title, will feel slightly

deceived. For it does not bring a general and comprehensive survey of *"sampling by variables"* and the variety of problems involved; the book mainly describes *"a particular system of sampling by variables"* to be used where an attribute sampling plan according to the tables of *Sampling Inspection* by the Statistical Research Group, Columbia University, or the Military Standard 105A is to be replaced by a variables sampling plan.

The basic pattern of the sampling plans described is that given by Wallis in Chapter 1 of the Statistical Research Group's *Techniques of Statistical Analysis.* The quality characteristic of the products inspected is supposed to be measurable on a variables basis; an upper or lower limit, U or L, is set, and we wish to control the percentage of items with a quality beyond one of these limits. To this end we measure a sample of n items, compute the average $\bar{x}$ and the standard deviation s from these measurements, and then require

$$\bar{x} + ks < U \text{ or } \bar{x} - ks > L, \tag{1}$$

as the case may be, k being a constant to be chosen in relation to practical requirements.

This single sampling technique is now supplemented with double sampling plans which are specified by three constants k_a, k_r, and k_t, and two sample sizes, n_1 and n_2. When after the first sample

$$\bar{x}_1 + k_a s_1 < U, \text{ the lot is accepted;} \tag{2}$$

and when

$$\bar{x}_1 + k_r s_1 > U, \text{ the lot is rejected.}$$

Since $k_r < k_a$ these conditions cannot be satisfied simultaneously. If neither of conditions (2) is satisfied, that is when

$$\bar{x}_1 + k_r s_1 < U < \bar{x}_1 + k_a s_1, \tag{3}$$

we proceed to take a second sample, the requirement being that ultimately

$$\bar{x}_t + k_t s_t < U \text{ for acceptance} \tag{4}$$

when $\bar{x}_t$ and s_t are average and standard deviation computed from the pooled measurements of first and second sample. Sequential sampling according to a similar principle would become complex and inconvenient in practice, and has not been included for that reason.

The above criteria hold if we assume the standard deviation σ of the lot unknown and variable from lot to lot. If past experience has shown that σ does not vary from lot to lot, known-sigma plans may be used by replacing s by σ in inequalities (1) to (4) and modifying the constants k.

Likewise a separate chapter is devoted to cases where we have to consider simultaneously an upper and a lower limit.

The choice of a sampling plan according to these principles is facilitated by an extensive set of tables and charts, which have been arranged after the

model set by *Sampling Inspection* of the Statistical Research Group and its offspring the Military Standard 105A.

After a choice has been made between three inspection levels, the plan is determined by lot-size and AQL-class. A full set of charts with operating characteristics of all sampling plans has been included, while additional tables are provided for finding the AOQL, or LTPD, and for deciding on tightened or reduced inspection.

The text accompanying these tables gives full instructions as to their meaning and application. One chapter of 17 pages dealing with the theory requires considerable mathematical knowledge for its understanding, but all other chapters have been written in a simple and clear style, while mathematical formulas have been largely avoided.

Yet despite the absence of formulas the text is mainly theoretical in concept, and many of the puzzles encountered in practice are not even mentioned as the following examples may illustrate.

The limits U or L, see above, are taken for granted, whereas it may often be an important preliminary step to decide by a statistical investigation whether they have been fixed in a reasonable manner.

It is repeatedly stressed that a first step is to decide what is the item to be inspected. In most cases, however, this is obvious enough, but it is more difficult to decide how the measurements are to be performed. For instance, in the case of metal sheets (pages 86–93) the thickness and Rockwell hardness may easily vary from the centre to the border of the sheets and if so it may be difficult to decide how and where the measurements have to be performed. Similarly special measures may sometimes be required to avoid systematic differences between different inspectors or apparatus. Such points have not been considered.

The sample size for known-sigma plans is about one-fourth that for unknown-sigma plans. One of the most useful functions of a control chart used in combination with sampling by variables would therefore be to indicate when sigma is sufficiently under control for the use of known-sigma plans; but chapter 9 provides no guide on this point.

In practice, lot sizes may vary considerably and sample sizes with them. But a control chart for varying sample sizes loses a great deal of its attractive simplicity and, therefore, of its technical importance. Hence if we are to derive full profit of the control chart technique in combination with sampling by variables a constant sample size may be of tremendous advantage. We should not forget the Lot-Plot technique, which has taught us the practical value of a constant sample size, and a simplified technique for computing the standard deviation.

On two points of methodology I disagree with the authors.

The standard deviation of a sample is defined by

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{(n - 1)}},$$

using $n-1$ in the denominator instead of n as is more common. This, I think, is a happy change; in more advanced applications of statistics the concept of the number of degrees of freedom is so fundamental that it should be introduced at an early stage, if we are not to be led into considerable confusion and contradiction later on. But the standard error of s as defined above is

$$\sigma_s = \sigma/\sqrt{2(n-1)}$$

and not $\sigma/\sqrt{2n}$ as stated on page 107; the factors B_3 and B_4 for control limits in table J should be corrected accordingly.

More serious objections can, I believe, be raised against the techniques recommended on pages 62 and 63 for finding the process average. Let us suppose that the critical limit U is 20, and that we receive 19 lots with average $\mu=17$ and standard deviation $\sigma=1$, and one lot with $\mu=25$, $\sigma=1$; 19 lots are completely good, 1 is completely bad, the process average is 5%. But if we pool these 20 lots into one grand lot, this grand lot will have an average $\mu=17.4$ and a standard deviation $\sigma=2.0$.

Should we consider this grand lot to be normally distributed, we would conclude that the process average per cent defective is 9.7%, a gross overestimate. It might be objected that nobody would ever consent to treat such a heterogeneous mixture of lots as one lot with a normal distribution, but this is exactly what is recommended on the pages cited. Samples drawn from successive lots are pooled into one grand sample, and the mean and standard deviation of this grand sample are treated as if they represent a grand lot with a normal distribution. This is only correct if the lots are under control, but then sampling inspection is not needed. Sampling inspection fundamentally applies where there is not sufficient control; in that case within-lot and between-lot variation should be clearly distinguished, and pooling of data from different lots is not permissible. The correct procedure is to estimate the percentage defective for each lot separately and average these percentages.

Summing up them, where Military Standard 105A or the Statistical Research Group's *Sampling Inspection* for attributes is to be replaced by sampling by variables, the reader will find in this book an immediate guide; while he who wishes to develop a system of his own, will find in it a great amount of information, which will be useful if combined with practical experience and ingenuity.

An Introduction to Statistical Science in Agriculture. *D. J. Finney*. New York: John Wiley and Sons, 1953. Pp. 179. $3.75.

H. W. Norton, *University of Illinois*

This little book has chapters entitled "The need for statistics," "Some problems of rates and frequencies," "Probability," "Properties and uses of distributions," "An experiment to compare two varieties," "The reduction of error," "Factorial design," "Sampling," and "Correlation and regression."

In addition, there is a preface, a valediction, references, and index. Its object is "to impart a knowledge of the general principles on which statistical science is founded and of the manner in which it enters into so many agricultural problems . . . the emphasis throughout is on the principles illustrated by the examples rather than on the arithmetic and algebra of calculation." I am one of that group of statisticians (mentioned by Finney) who think the correlation coefficient is now of very trifling importance, especially in an introductory book. This topic can hardly have much of the merit of familiarity attributed to it by the author, and the space devoted to it might better have been used to give a simple example of analysis of covariance.

A number of formulas are given and used arithmetically, but there is no algebra nor other mathematics with the sole exception of the derivation of a formula for the sample number required to realize a specific confidence interval. Type, paper, and binding are good and there are few typographical errors.

It is a curious fact that there appears to be a real need for nonmathematical books on statistics. That is, it is curious that there are many people who are established scientists, and many others offering to become scientists, who have so little mathematics (in spite of a great need thereof) as to warrant the preparation of severely nonmathematical books expounding an essentially mathematical subject. It is reassuring in a way, in the midst of current attacks on education in the public schools, accompanied sometimes by assertions that standards are higher elsewhere, to note that this book is based on (and presumably is the substance of) a "course of eight lectures . . . to undergraduates reading agriculture at Oxford." I have long held the view that the research scientist's most urgent statistical need is for an appreciation of the essential ideas of the design and analysis of experiments, and that those ideas can be appreciated without recourse to mathematics. Therefore I expect Finney's book will probably have wide appeal, and it seems well suited to its purpose of conveying appreciation of principles.

An important feature of the book is its stress on the relation between the research worker and the statistician, and the continual emphasis on the desirability of expert statistical advice at all stages of the work but especially before experimentation is begun. This culminates in the "Valediction," with its twelve rules of respectable statistical conduct for research workers. The first is "When you propose to undertake an experiment or a sampling investigation of a kind that presents any novelty to you, consult a statistician at an early stage of your planning." This will serve as an example while enabling me to point out my only substantial criticism of these rules, that the phrase "of a kind that presents any novelty to you" should have been omitted on the ground that recognition of relevant novelty will often require consideration by an experienced statistician. I was glad to find (p. 126) Finney's observation, "even a year . . . is not an excessive margin, for consideration

of the relative merits of different designs takes a long time and is better done in short spells over a period than in a single concentrated effort." How often the statistician has occasion to agree!

There are a few statements to which exception must be taken, and which do not appear to be merely the result of such simplification (usually called oversimplification) as reasonably would be expected in such a book. The assertion (p. 25) that Yates' "correction is always $\frac{1}{2}$; and (p. 33) "applies *only* to 2×2 tables" may be defensible in a chapter which emphasizes contingency tables, but is likely to prove confusing to the novice who looks into more than this one book. The statement (p. 156) that the covariance "must always be intermediate in magnitude between" the two variances is wrong: it may not be larger in magnitude than the square root of the product of the variances. The statement (p. 102) that "both the smaller dressing of sulphate of ammonia and the ammonium humate gave yields significantly higher than that without nitrogen" may prove disconcerting to the careful reader, because the differences mentioned are exactly at the 5% point (following Finney's practices in rounding) for the comparison of yields in pounds per plot, but do not quite reach the 5% point for tons per acre, and retention of additional decimals leads to the conclusion that significance is reached only by accumulation of rounding errors. Occasionally the choice of words is inappropriate as (p. 75) the use of "indubitable" to describe the superiority of one variety to another after their difference of yield has proved significant. Some statements would be improved by minor changes, such as (p. 60) "conditions that are as alike as possible," which is likely at once to coincide with the readers' prejudice and to miss an opportunity to emphasize the potential importance of a proper distribution of effort in experimentation, would better say "as practicable" or "as reasonably possible," and (p. 68) "the difference is greater than would occur by chance" needs to say "would be at all likely," so as to keep always before the novice the idea that *any* difference may have occurred purely by chance. I could not verify the probability near the top of p. 57, nor the fiducial limits on p. 159. There is a misprint in table 6.1, where the mean for block *V* should be 19.0; on page 98, where the divisor in the formula for *F* should be 3.0; and on p. 168 the references are to table 9.3.

Agricultural Prices. *Frederick Lundy Thomsen* and *Richard Jay Foote.* New York: McGraw-Hill Book Company, 1952. Pp. xi, 509. $6.50.

L. J. Norton, *University of Illinois*

This is a thorough revision of a 1936 book by the senior author. It is a text book intended for students principally in agricultural colleges. It bears the imprint of extended experience in teaching, government economic

research, and practical experience in business. It is divided into three parts:

1. A review of simple economic principles, discussion of some materials, presentation of the elements of economic fluctuations, and an analysis of government price programs under the over-all heading of "price determination and discovery."

2. Price analysis and forecasting.

3. A series of commodity reviews, labeled "commodity prices." The emphasis is on simple, direct tools rather than on complex mathematical methods.

Economic statisticians, particularly those engaged in commodity analysis, will find Section II of particular value. The chapters on commodity prices are not up to the standard of those in the first two sections in getting the problems into clear focus.

The authors hold a hard-headed point of view with reference to the value of overly elaborate analysis in forecasting. Their point of view might be summarized: Get all the information you can out of the best analysis you know how to make. Test your results with all the logic and common sense you have. Apply them with judgment and be on the alert for facts not used in your basic analysis. From his observations on forecasters and forecasts this reviewer deems this to be a highly sensible point of view. The authors suggest that pin-point forecasts are not needed but rather "a degree of interpretation which will enable the user to anticipate the direction and to some degree the extent of the movement" (p. 351).

In discussing price determination both the aggregate approach and the older individual commodity approach are used. No time is spent on hair-splitting differences or too-elaborate types of analysis. The students should get the points. Perhaps the terms supply and demand are used too broadly and so lack precision, but this may be justified on the grounds of teachability. The discussion of the "relation between cash and future prices" is not up to the standard of the book. Quite possibly many traders in the future's markets have better bases for formulating expectations than this treatment would lead one to believe. In some form or other many of them have grasped some of the ideas which the authors later develop as to reasons for price trends. In discussing the general price level too much emphasis is put on banking and credit. Perhaps the authors do not infer causation but their treatment would leave such an impression in the reader's mind. After a period when monetary inflations, devaluations, and many other developments have affected the levels of prices in various parts of the world, their analysis seems oversimplified. This may be important because in connection with the general price level the price analyst may find an important element which he may have omitted from his statistical calculations. In this area the authors do not seem to have done work in depth as in other phases of price analysis.

On the whole this is a good job and both students and workers in the field of price analysis will find it valuable.

A Textbook of Econometrics. *Lawrence R. Klein.* Evanston, Illinois, and White Plains, New York: Row, Peterson and Company, 1953. Pp. ix, 355.

KENNETH J. ARROW, *Stanford University*

THE appearance of this book marks an important stage in the development of econometrics, since for the first time there is available a useful textbook. To say that Klein's work is the best of its kind would be correct but very inadequate, since he has virtually no competition. Some other books, which have appeared under similar titles, are very useful in their own right, but they do not meet the need of a beginning course in econometrics (as opposed to mathematical statistics) presupposing a reasonable but not excessive knowledge of mathematics as well, of course, as a good knowledge of economics.

Econometricians are very fortunate indeed that this text, which will undoubtedly be the standard one for many years, is so truly excellent. Klein's preeminence as a practitioner of the econometric art is well known, but one would not necessarily expect such an extraordinary level of didactic skill from a mathematically-inclined economist (although the readers of *The Keynesian Revolution* might not be surprised).

After an introductory chapter on the econometric approach, there is a forty-page summary of the basic principles of mathematical statistics. While about as good a job has been done as is possible in the space,[1] I seriously doubt the usefulness of the attempt. The material the author seeks to cover requires, I would judge, about one semester to master, and no attempt to speed up the learning process is apt to be successful. I think it preferable to require a basic course in mathematical statistics as a prerequisite.

Chapter III, entitled, "Estimation of Aggregative Models," is in many respects the core of the book. A simplified macro-economic model is developed, and the method of least squares is then introduced as the maximum-likelihood method of estimating one of the equations. In more general cases, it is noted that the problem of identification arises; the concept is rigorously defined and its application to linear models stated. The estimation of an aggregative model by the method of maximum likelihood applied to the whole system is then considered, with attention paid to the special cases where the statistical method has a simple form. The method of instrumental variables and the limited information method are next introduced, though the explicit derivation of the formulas from maximum likelihood considerations is omitted. The chapter concludes with a discussion of confidence intervals for the estimated parameters.

The next chapter is a detailed description of the computational designs of the least squares, maximum likelihood full-information (under the assumption of a diagonal covariance matrix of the disturbances), and maximum likelihood limited-information methods. The exposition is first-rate; for a student, it could hardly be improved upon.

[1] Some minor errors in this chapter have been noted in the review by R. Solow, *American Economic Review*, Volume 43 (1953), 947–50.

the measurement of income or wealth, with the meaning of the concepts measured and with the interpretation or analysis of the measures. Most economists and statisticians engaged in studies of income and wealth are concerned with basic problems of measurement, the formulation of concepts and definitions, the accumulation of data from diverse sources ranging from administrative statistics to family living surveys and the adaptation of these data to the conceptual frame-work of a particular system of accounts. The greater number of entries in Volume II, as in Volume I, present national estimates or survey data on income, wealth, and related topics for various dates.

In general, the estimates of national aggregates seem to be conditioned in all countries by the same limitations imposed by the nature of the source material, and the supply of data, and by the financing of the necessary research. The first of these conditions is apparently leading towards some uniformity in the details of the national accounts for various countries under the influence of the various recommendations of the United Nations' Statistical Commission. The second explains the variation in the volume of estimates among countries in possession of the source data and the technical skill required for their utilization in the construction of systems of national accounts.

The discussions of concept in the writings listed in Volume II display, on the other hand, no clear tendencies towards the resolution of the controversies that have filled the literature on national accounts for the past two decades or more. From the viewpoint of a statistician with a bias towards empirical methods the controversial issues can not be settled until they are submitted to a thorough logical analysis. The operational nature of the basic concepts needs clarification; the limitations on measurement should be recognized; and the distinctions between the concept specified for measurement and the theoretical construct of the same name should be sharply drawn. Until the logic of this field of measurement has been considerably refined, the contribution of statisticians to the improvement of the quality of the primary estimates will be seriously restricted. A few entries in Volume II discuss the potential usefulness of sample surveys for the collection of data needed for the purposes of national accounts and the need for developing measures of the error in estimates. The statistician's skills can not, however, be effectively employed until the basic concepts are made to correspond more simply with magnitudes that can be directly observed than is now the case.

Basically the concepts specified in national accounts are synonymous with the operations for determining the particular measurements. National income analysts have adopted the tools of the accountant and are developing rules for their application to data from diverse sources relating to the economic activities of the productive units in the particular country. The productive units fall into three classes, business enterprises, households, and governments. The economic activities of the three classes of units are not distinct

and they overlap in different ways and in varying degrees among the nations now attempting to accumulate national income estimates and other social accounts. The procedures of accounting apply directly to one class of units, business enterprises and such concepts as wealth, income, production, consumption, and capital formation can be differentiated in the records of the economic activities of these units. More or less arbitrary but fairly standardized rules have been devised for fitting such activities of other classes, households and governments, into the same structure of measurements as overlap with or substitute for the activities of business enterprises. Thus the food the farm family consumes that otherwise might be sold by the farm business; the home occupied that might otherwise be rented by the owner as a landlord; the goods and services provided by a public authority that elsewhere are produced by business enterprises can be described by essentially the same structure of accounting concepts.

Difficulties arise when operational definitions are extended well beyond the domain of experience with their application. The Bibliographies indicate clearly that the scholars engaged in the measurement of national accounts were, as of 1949, still unwilling to cope with the fundamental limitations on the extrapolation of concepts beyond the region of their applicability. The blurring of concepts outside of the original area of definition which occurs in all fields of measurement is particularly well illustrated by the problems encountered in describing the economic activities of "subsistence" units for the purposes of national accounts. The differentiation between production and consumption, saving, wealth, income, and similar concepts cannot be simply established by adopting the operational definitions that serve for the aggregation of the accounts of business enterprises. Volume II of the Bibliography, like Volume I, lists very few references to any attempts at measurement of the activities of subsistence units. The small number of entries does not indicate, however, that the analysts have rejected this domain as not susceptible of measurement in terms of prevailing concepts. There are as many references which list the failure to measure production for home consumption as an element affecting the accuracy of comparisons of national income estimates among countries and between dates. The absence of "data," not a consideration of the validity of concepts in this domain, probably explains the small amount of research on this subject.

The case is quite the contrary with government purchases which, beyond the area of activities overlapping with business enterprises, present the difficulties of extrapolation to another scale. Government accounts and balance sheets are not easily summarized to yield the concepts specified for aggregated accounts of a whole economy. There are a fairly large number of direct references, some twenty papers dealing explicitly with the introduction of government activities into national accounts and nearly half of the general papers on concept treat the problem. There are doubtless many ways of explaining the apparent concentration of analysts in the years 1948–49 on the problems involved in fitting government activities into the structure of

national accounts and answering Kuznets' questions in the preface to the Bibliography, Volume I, "Why do 'official' estimators in industrially developed countries in recent years adhere so closely to the treatment of all government purchases as final product, whereas unofficial estimators do not easily accept the underlying assumption? Does the former treatment permit greater ease in planning and budgeting governmental activities and tracing their impact on the short-term economic situation of the country?"

Data in the form of government accounts do exist in most countries and it is not surprising to find the empirical workers concentrating on the problems of extension and adjustment of concepts to this sector of economic activities. The answers to Kuznets' two questions will be given by experience with statistical analysis of the national estimates that have been accumulating in volume sufficient to warrant this Bibliography to cover the output of two years.

The estimates of national product, national income, and other forms of social accounts have become exceedingly important in national policy, planning and forecasting. The tools of the statistician may not, for some time to come, be of much use in the primary processes of data collation and estimation but in the processes of analysis the statistician has much to offer. The real refinements of the concepts in national accounts can be expected to come as experience with statistical analysis of the existing estimates, however incomplete, inaccurate, and inconsistent they may be, reveals the structure of definitions that is practically most useful for the purposes the estimates are intended to serve.

The relative virtues and values of the sundry concepts for prognosis, forecasting and general economic studies will probably be decided by the results of statistical analysis long before the scholars can agree on the identification of the concepts specifying the measurements with theoretical abstractions. The accumulation of observations and methods of empirical analysis may soon render some of the controversial issues obsolete, even meaningless. The stress, for example, on national income or some other aggregate as an absolute measure of "welfare," a really vague abstraction, in the formulation of concepts will doubtless disappear as experience with analysis is extended in time and over the nations of the world. Analysis to date, it is true, can produce little that merits such optimism about the progress of empirical knowledge of the numerous economic magnitudes that fill in the national accounts. In particular, the much debated "consumption-saving function," which accounts for at least 10 papers listed in the Bibliography, Volume II, does not seem to have produced any promising new tools of analysis. Nevertheless, the increasing volume of estimates of national income, its variants and components, national wealth and social accounts has already stimulated new work in time series analysis and surely will lead to some new developments in the utilization in economic analysis of observations on geographic areas. The estimates recorded in the Bibliography for 1948–49 cover more than 70 countries or colonies for varying dates or ranges of dates. Surely this

growing wealth of information, conveniently located through the Bibliographies, will inspire more than a few statistical analysts to surmount the difficulties imposed by variability in concept and in the precision of estimates and devise the means of analysis. Empirical science thrives on the continuous revision of theoretical concepts imposed by experimental findings and the steady refinement of methods of measurement and analysis that result from these revisions.

The Role of Federal Crédit Aids in Residential Construction. *Leo Grebler.* New York: National Bureau of Economic Research, Inc., 1953. Occasional Paper No. 39. Pp. 76. $1.00. Paper.

SHERMAN J. MAISEL, *University of California (Berkeley)*

THIS paper is a concise review of the direct impact of the Federal government on residential financing. Data have been gathered from a wide variey of sources. They have been well organized so as to give to the reader a clear picture of the extent of Federal intervention in the real estate financing field. By judicious selection and analysis, the author clearly illuminates the possible results of these governmental policies. While the analysis is clear, lack of data has made it impossible to appraise fully most of the hypotheses concerning effects that are presented. This lack of data, together with the large amount of piecing together of disparate statistics required for this study, is a clear indication of the failure of the Federal housing agencies to collect and publish adequate statistics required for an accurate evaluation of their programs and policies.

The paper points out that under the programs of the Federal Housing Administration and Veterans Administration, the government in recent years has insured or guaranteed between 40 and 50 per cent of loans made on new housing. As a result of these governmental policies, the level of house construction was probably somewhat higher than it would otherwise have been; the market for houses was almost certainly widened; and the costs of ownership may have been lowered. Although primarily as by-product results of actions taken in response to other needs, there have been important impacts on the type of houses built and on institutional practices in the lending markets.

Many readers will find that the value of this review of policy is increased by a supplemental statement of the author as to the conclusions concerning the future that he draws from the past trends. By adding such a statement, the author obviously opens himself to attack from those who disagree with his reading of history. This reviewer feels, however, that the clear statement of future policy problems which may arise from the action taken to date greatly enhances the value of this paper. Providing, as in this case, sufficient detail is given to make disagreement possible, a consideration of future problems by an author completing a detailed analysis of the past is a valuable addition to the study, whether or not one agrees with the conclusions.

Retail Prices and the Consumer Preference. Studies in Business and Economics, Volume 6, Number 1. University of Maryland, College Park, Md.: Bureau of Business and Economic Research, 1952. Pp. 8.

SOPHIA GOGEK, *Standard Oil Company (N. J.)*

THIS pamphlet contains three brief papers: "Price Variations in Men's Clothing by Kind and Location of Store," "Retail Prices, Income and Race," and "Consumer Patronage of Independents and Chains in 59 Cities," each consisting primarily of a summary of the data analyzed. The first compares the prices at neighborhood and centrally located specialty and department stores within and among eighteen cities of eleven items of men's clothing. The second investigates the relative prices in one negro and three white (low, medium, and high rental) neighborhoods in Chicago of a selection of dry goods, men's clothing, and women's clothing. The third summarizes consumer preferences, in shopping for food and for drugs, as between independent and chain stores, and as between neighborhood and centrally located outlets.

Analyses of this type would be both more interesting and more useful if there were more discussion of statistical problems and of the economic and cultural factors responsible for the behavior of the data as recorded. (The second paper, which does go into these matters to some extent, is much the best of the three.) A listing of the sources of data, where they are apparently secondary, would also be helpful.

Industrial Specifications. *E. H. MacNiece.* New York: John Wiley and Sons, 1953. Pp. xiii, 158. $4.50.

J. H. CURTISS, *New York University*

THIS little book gives a description of several different types of written industrial specifications now in use. The classes discussed are raw-material specifications, process specifications, product specifications, and purchase specifications promulgated by public agencies. The exposition is illustrated rather extensively by examples.

The treatment is generally descriptive rather than analytical, and accepts current practices without criticism, constructive or otherwise. The suggestions as to the techniques of preparing specifications are largely confined to listing the topics which should be covered and to urging that clear language be employed.

This serves a useful purpose, and the book should be an influence toward better specifications. On the other hand, the deeper problems of specification-writing are dealt with only briefly and qualitatively. The problems which the reviewer has in mind are exemplified by these: Should acceptance sampling requirements be included in the written specification along with design requirements or should they be issued separately? Should design requirements ever be stated in terms of the behavior of random samples? (In several of MacNiece's examples, this does occur.) How should *laboratory research* data be

processed so as to derive valid tolerance limits? Are inspection data better than laboratory data for this purpose? And so forth. MacNiece handles problems of this sort largely by generalities, such as "Quality control and industrial specifications are blood brothers, and each depends on the other."

But the choice of subject matter is always an author's prerogative. The fact that this little volume is non-technical makes it easy to read, and MacNiece writes fluently and well. Quite apart from its intended use, the book should provide some useful case-history material for teachers of quality control and industrial management.

Analysis in Dental Research. *Neal W. Chilton.* Washington, D. C.: Office Technical Services, U. S. Department of Commerce, 1953. Pp. 216.

GEOFFREY BEALL, *University of Connecticut*

THIS introductory text on statistics is devoted entirely to data obtained in dental studies. One's first impression is to regret that not only must we accept the modern departmentalization of knowledge but must suppose modern man incapable of transferring mental operations from one field of fact to another. On second thought, the dental student, research worker, and clinician will indeed enjoy a book where statistics is so richly illustrated in their terms. The book bears the marks of having been tried and proven in class; the discussion is pleasant and reasonable.

The general treatment shows the mark of being unduly influenced by the extensive and varied literature of texts introducing statistics. Thus, in connection with the calculation of a standard deviation from a sample, the author follows the very old and curious procedure of recommending a biased estimate, having divisor N, when N is greater than 30. On the other hand, as in connection with concepts of chance variation and the binomial distribution he is influenced by the vogue of the moment which would circumscribe statistics to an elaboration of combinatorial probability. This confusion of position does not help general organization. Thus correlation is presented in an infelicitous manner involving the estimation of standard deviations and means. It should, in the present context, be reduced to a form involving only sums of various kinds. An unhappy mixture occurs in the discussion on frequency graphs together with the general nature of numbers and the presentation of time series.

Various particular faults may be found. Thus the discussion of the mode is confusing since the term is not used in the ordinary statistical sense but here means the most common class. The author gives a lengthy and involved discussion of "normal curve analysis of the four-fold table" without apparently realizing that he has calculated a value of chi (not squared). He then goes on to presentation of the contingency for the 2×2 table to get chi-squared. In neither case does he include Yates' correction for discontinuity.

In spite of faults the book still has practical virtues. It concerns itself first with the analysis of qualitative data and then with quantitative data as a less

common even if more happy situation. The author does not stress the t-distribution but considers tests of various kinds on a simple normal basis, although he devotes a terminal chapter to a very satisfactory discussion of the t-test and analysis of variance.

Intrastate Migration in Michigan: 1935–1940. *Amos H. Hawley.* Michigan Governmental Studies No. 25: Ann Arbor: University of Michigan Press, 1953. Pp. 199. $1.50. Paper.

JOHN FOLGER, *Southern Regional Education Board*

THIS is a book of facts about migrants in Michigan. As an example of pure empiricism it is outstanding, description hardly ever giving way to analysis or interpretation. A good index of the descriptive nature of the work is the fact that it contains 106 tables, 28 figures, and only one reference to any other study in the field of migration.

Hawley is to be commended for the workmanlike job of reducing the tremendous amount of detailed data contained in the subregional tabulations of migration to understandable proportions. But the study never goes beyond the initial framework provided in the Census data. No interpretation is given either by comparison with migration in other times or places or in terms of conditions or changes in the subregion of origin or destination. The student of population seeking new theories of migration or new tests of old hypotheses will be disappointed. The facts are there, and all of the pleasure and hard work of assessing their meaning is left to the reader.

The measures of migration used are all easily understood, proportions and simple rates being employed throughout most of the book. Greater use of more refined rates and indirect standardization techniques would have been valuable in removing the effects of differences in characteristics between migrants and nonmigrants when studying the differences in other characteristics. For example, how much of the difference between migrants and nonmigrants in marital status is attributable to age differences between the migrants and the total population? While adjusted rates and proportions are harder for the reader to comprehend, they are almost a necessity in the process of untangling the complex of variables associated with the process of migration.

It is this reviewer's hope that future studies of migration utilizing the rich detail of the subregional tabulations will be able to perform both the competent summarization of the data that Hawley has provided, and a systematic testing of hypotheses about migration and the characteristics of migrants.

Sample Surveys of Current Interest. (Fourth Report.) Statistical Papers, Series C, No. 5. New York: Department of Economic Affairs, Statistical Office of the United Nations, March 1952. Pp. 56. Paper. 50 cents.

HAROLD NISSELSON, *Bureau of the Census*

THIS Report is the fourth and latest in a more or less annual series begun in 1948 at the suggestion of the Sub-Commission on Statistical Sampling

of the United Nations Statistical Commission, "to assist those interested in the application of modern sampling techniques by making available, in summary form, the experiences gained by various statisticians and statistical organizations in the field." Altogether, 33 of the 60 U.N. countries are represented in one way or another in the four Reports published to date. Perhaps not unexpectedly, there have been no reports from Russia nor from any of the "Iron Curtain" countries except Czechoslovakia (twice), Hungary (twice) and Poland (once). The projects reported are presumably not to be taken as a proper sample of all sample surveys being carried out in the 60 U.N. countries. However, the general impression left by the Reports is that the ideas and techniques of modern probability sampling have achieved a rapid and wide circulation, both geographically and in subject matter dealt with. On the other hand, hardly any reference is made to non-sampling errors in surveys. It is clear that elsewhere, as in the U.S.A., the general theory of survey design—in contrast to its sampling phases—is in a relatively undeveloped state.

The information for these Reports is obtained by circularizing the national statistical offices of all U.N. countries (in the U.S., the Division of Statistical Standards, Bureau of the Budget, Executive Office of the President), with dependence on the cooperation of those offices for replies. The reports, there fore, have three characteristics. First, they appear to include only work of official and quasi-official (e.g., The Indian Statistical Institute) organizations. Second, not all current projects are reported, as is explicitly noted for India, the U.K., and the U.S.A., and the selections may be considered somewhat arbitrary. Third, the project descriptions do not completely adhere to the recommendations of the Sub-Commission itself.[1] Information such as variances of sampling units or unit costs is scanty; and the summaries tend to be long on description and short on evaluation. (The U.S.A. is not outstanding in these respects.) Statisticians interested in technical details of the design of a given project, or in a realistic evaluation of the experience in carrying the project out, will therefore generally need to write to the United Nations Statistical Office or the sponsoring organization for further information. Nevertheless, the Reports do provide a compilation in a single place of some of the more important official sampling projects being carried on in various countries, many of which have not been published as papers on sampling. For example, the present Report contains the only published statement currently available on the applications of sampling in the enumeration and processing of the 1950 Censuses of Population and Housing. For this purpose the Reports are more comprehensive, and should be more useful, than specialized listings presently appearing in scattered journals. It presumably lies with statisticians themselves to decide how useful this series may be; and, by their cooperation to remedy any short-comings in the Reports.

[1] "The Preparation of Sampling Survey Reports," Statistical Papers, Series C, No. 1, Statistical Office of the United Nations.

Age and Achievement. *Harvey C. Lehman.* Princeton, N. J.: Princeton University Press, 1953. Pp. xiii, 359. $7.50.

LEONA E. TYLER, *University of Oregon*

Too often in psychology research workers have made generalizations that extend far beyond the evidence their data present. This book is a striking exception. Its 331 pages with their 170 graphs and 61 tables are all related in some way to the important question, "What are man's most creative years?"

What the author has done is to analyze facts reported in standard reference books with regard to the ages at which great men have made their outstanding creative contributions. Science, philosophy, music, art, and literature have all come under his scrutiny. The result is a large number of curves, each of which shows the average number of important contributions for each five-year age period in the area of human achievement to which it applies. All of these curves are placed on the same scale by the simple device of reducing the figures to percentages. Whatever its absolute value, the highest number, representing the peak age interval, becomes 100 per cent and the others take lesser values accordingly. Other statistical information is given in extensive tables so that the reader can quite easily locate numbers of cases, medians, means, and standard deviations for the various distributions. The discussion emphasizes the modal values, the peaks of the curves, more than anything else, but the other figures are not ignored. The derivation of basic data from biographical dictionaries insures freedom from bias since such figures were compiled by men who were not concerned at all with this particular question.

In addition to his discussions of the central problem, the author includes some chapters on supplementary issues, such as the relationship between early achievement and total output, and comparisons of age trends in earlier and later periods of history.

Several conclusions are drawn from the data presented. First of all, the vast majority of the curves show a remarkable similarity. The peak occurs early, most commonly in the 30's, but as low as the 26–30 interval for chemistry and as high as the 40–44 interval for light opera and musical comedy, for "best books," and for metaphysics. All curves show a gradual decline, with some achievement indicated even into advanced ages. It is perhaps noteworthy that the data for sports and for such activities as chess show the same trends as the others. Championships are won most frequently by persons under 36.

Secondly, when the works agreed by most critics to be of the very *highest* quality are singled out for special attention, the curves drawn from these figures show earlier peaks and a more rapid descent than do those for the less selective distributions. This finding is stressed again and again. The more *quality* as distinguished from quantity is considered, the younger is the average age at which it is achieved.

Thirdly, data for incomes and for positions of leadership produce curves

which are quite different from those for creative achievement. Their peaks occur most commonly at ages of 50 and higher.

The author does not present any one type of explanation for these findings. He reminds us periodically that these are not biological curves representing the rise and decline of vigor or intelligence in any *one* individual, although if there were such developmental trends they would reflect themselves in this way. The extensive documentation in Chapters 13 and 14 with regard to outstanding contributions made at very early and at very advanced ages, as well as the evidence that trends have changed somewhat from one historical period to another, argue against a *purely* biological explanation. There are a number of other possible causes centering around motivation and around the circumstances in which great men typically find themselves at different periods of life. These are listed in detail in the last chapter.

It is possible to criticize Lehman's report on two counts, one rather superficial, the other less so. First, it would have made better reading if the author had worked over and reorganized his material instead of simply combining into a book chapters that had already been published as journal papers. There is a great deal of repetition from chapter to chapter, as things stand, and the same points are made again and again. Second, the statistical work is confined entirely to *descriptive* statistics. Our judgments must be based on a scrutiny of the figures and tables, without benefit of significance tests of any kind. So far as the main question is concerned, this is probably satisfactory, since the position of the "peaks" is easily determined by inspection. But other questions would seem to call for tests of the statistical significance of differences. How similar are various curves in shape? Is there really a tendency toward bimodality as many of the figures in the first few chapters would suggest? Is the change from an earlier to a later period, as discussed in Chapter 18, significant? It is unfortunate that Lehman's interest in the *shape* of the age curve has in some instances led him to give us data to which significance tests cannot very well be applied. For instance, in order to show peaks clearly, five-year intervals were selected in different ways for different bodies of data. Thus for German lyrics and ballads, the intervals are 17–21, 22–26, 27–31, etc. instead of the 20–24, 25–29, 30–34, etc. which are used in most of the tabulations. Distributions set up in this way with intervals chosen after inspection would not be statistically comparable. Another instance occurs in Chapter 18 where the conclusions about historical periods would be strengthened by significance tests. What Lehman has done in each of his groups is to compare curves for the 50 per cent born earlier and the 50 per cent born later. The result is that the actual historical period represented varies from one comparison to the next. For example, a geologist born in 1800 is included in the "early" group, whereas a mathematician born in the same year appears in the "later" group. Thus different groups of scientists could not very well be combined.

These criticisms, it is true, apply to only a fraction of the data presented. Other workers wishing to test various hypotheses about age trends in achieve-

ment will find this book a mine of information. Much use can probably be made of means and standard deviations as given in the various tables. While we can wish that the author had gone further in using this large body of data to enlighten us as to *why* these trends appear, we can at the same time be grateful to him for digging out the figures which show the trends themselves so plainly.

Backgrounds of Human Fertility in Puerto Rico: A Sociological Survey. *Paul K. Hatt.* Princeton: Princeton University Press, 1952. Pp. xxiv, 512. $5.00. Paper.

HENRY S. SHRYOCK, JR., *Bureau of the Census*

THIS book is the first and most general report of a field survey conducted in 1947 under the auspices of the Office of Population Research, Princeton University, and the Social Science Research Center, The University of Puerto Rico. The study was designed to "throw light upon those basic attitude patterns and life conditions which affect fertility levels in Puerto Rico." Its conceptual design has much in common with the 1941 Indianapolis Study of Social and Psychological Factors Affecting Fertility.

The probability sample used by the Insular Bureau of Labor Statistics in its monthly survey of the labor force provided a prelisting from which the present sample was drawn. Of all listed households, 92 per cent were located and at least one adult interviewed. Within these households 92 per cent of the adults were interviewed so that about 85 per cent of the adults covered by the original list were probably interviewed. The actual sample was checked for representativeness against the 1940 Census and by comparing households missed with neighboring households. There were moderate biases in the direction of the exclusion of youths and the economic extremes. (Data from the 1950 Census, which have subsequently become available, indicate somewhat less bias in the sample with respect to the sex and age of adults.)

The supervisors of the field work were social science graduates. The 65 experienced interviewers, all of them women, were given two days' training plus a conference after two practice interviews. The schedule was pretested upon Puerto Rican women living in New York. The final forms, as well as the instructions to the enumerators, are reproduced in appendices, although only in the original Spanish. This rather long and intensive schedule consisted of several parts: form 1 applying to each household, form 2 to each person 15 years old and over, and forms 3 and 4 to each person who had had any kind of marital experience (legal, consensual, or concubinal). On the other hand, the 1950 Census question on children ever born was asked of all women, including the single; and the over-all nonresponse rate was quite low.

After an initial chapter on "Methodology," the next four chapters are entitled, "The General Patterns of Social Conditions and Social Attitudes," "Socio-economic Status: Rental Value and Educational Level," "Rural-

urban: Birth and Residence," and "Age Differences," respectively. These chapters treat the relationships among various background factors but do not deal with measures of actual fertility. The relationships are examined by tetrachoric correlation, with partial and multiple as well as zero order coefficients being presented. In many instances there would seem to have been enough cases to permit further cross-classification, which would have been more satisfactory than holding a factor "constant" through partial correlation.

The factors brought into this correlation analysis include attitudes toward a number of social questions, including, among others, the desirability of consensual marriage for both men and women, ideal age at marriage, and ideal size of completed family. The attitude measures themselves are examined in several interesting ways (for example, ideal age at marriage is compared with own actual age at marriage and number of children desired for daughter is compared with the woman's own actual fertility). The attitude questions and the construction of the measures based on them seem relatively primitive after the reports of the Research Branch, Information and Education Division, War Department, although it must be allowed that little had been published by Stouffer and his associates on their methodology by the time the present study was designed. (The reviewer suspects that the rather disappointing findings of the Indianapolis Study concerning the effects of psychological factors on fertility would be replaced by much more positive results could this new apparatus of attitude research be employed.)

Tetrachoric correlation is also employed extensively in the chapter on "Differential Fertility," which contains much of the meat of the study. It seems unfortunate that, for mechanical considerations of processing the data, fertility is expressed in terms of pregnancies rather than of live births, since the latter were probably reported more accurately. Robert Osborn, Jr., has contributed a long chapter on "The Trend in Fertility," using mostly cohort analysis of the survey data but also pertinent census data and vital statistics. The downward trend that he is trying to measure seems to be a recent and slight one—less pronounced, for example, than the downward trend of fertility in the British Caribbean possessions. This kind of job needs to be done, but it is often thankless work because many readers are frustrated by the absence of definite conclusions as to the size of the various suspected biases and their net effect. Osborn is dealing with data that just do not yield such conclusive answers, however.

In conclusion, this book represents a sound job by the late Professor Hatt and his associates. Their findings are interesting and significant. Insular attitudes toward such things as family limitation, ideal size of family, ideal age at marriage, education wanted for children, and the employment of women outside the home are surprisingly like those on the mainland but also greatly at variance with current practices in Puerto Rico. If only a part of these changing attitudes can be translated by the oncoming generation into

actual conduct, there would be a very substantial drop in the high fertility levels now prevalent. From the standpoint of statistical methodology, this study contains few innovations; but the report does demonstrate that a fairly complicated kind of sample survey can be successfully carried out in this kind of underdeveloped area.

An Approach to Measuring Results in Social Work. *David G. French.* New York: Columbia University Press, 1952. Pp. xiv, 178. $3.00.

JOHN E. WALSH, *U. S. Naval Ordnance Test Station*

THIS book is a report on the 1950 Michigan Reconnaissance Study of Evaluative Research in Social Work, which was sponsored by the Michigan Welfare League and financially supported by the James Foster Foundation. "Evaluative research" refers to research performed to evaluate a practice or a policy. The book is not itself a piece of evaluative research but rather a survey of the possibilities in applying evaluative research to social work. Although carried out with reference to social work in Michigan, the analysis and recommendations are believed to have general applicability. The presentation is both logical and lucid.

Three general procedures were used in performing the Reconnaissance Study. First, factual information which appeared to have application to the planning of a research program in social work was assembled for the field of research and for the field of social work. Second, the available literature dealing with the general field of applied research, with emphasis on social science research and social work, was reviewed. Finally, both individual and group conferences were held to supplement the available printed information. These conferences were used to obtain suggestions and observations as well as to test portions of the analysis and conclusions developed during the study. As the study progressed, the emphasis shifted from evaluation to research. This was due to the great difficulties involved in performing valid and useful research in the social work field. However, the evaluation aspect was retained as a meeting place for the practical concerns of the social worker and the more theoretical attitude of the researcher.

An outline of the content of the seven chapters of the book furnishes a good indication of the approach used in the Reconnaissance Study. Chapter I describes the general motivation which led to the study. Chapter II considers the situation in Michigan and emphasizes the large financial investment in social work (approximately $125,000,000 in Michigan for 1950). Chapter III states some general types of questions which professional and lay persons raise concerning social work and studies these questions from the viewpoint of planning a research program. Chapter IV analyzes what is involved in doing evaluative research. Four representative evaluation studies in social work were reviewed to determine their strong and weak points. Chapter V considers the relationship between theoretical and applied re-

search as a basis for planning a research program in social work. Chapter VI states eight criteria which are believed to be important for any continuing program of research in social work and outlines an administrative setting which appears capable of satisfying all the criteria. Chapter VII extends the general considerations of the preceding chapter into some moderately specific suggestions concerning the establishment of an institute for research in social work.

The four methodology analyses of evaluation studies considered in Chapter IV are themselves of interest. Only the study title, the authors, and the person making the analysis are stated here: (1) "Measuring Results in Social Casework: A Manual on Judging Movement," by J. McVicker Hunt and Leonard S. Kogan with analysis by John G. Hill. (2) "Changing Attitudes Through Social Contact," by Leon Festinger and Harold Kelley with analysis by Leon Festinger. (3) "An Experiment in the Prevention of Delinquency," by Edwin Powers and Helen L. Witmer with analysis by Helen L. Witmer. (4) "Unraveling Juvenile Delinquency," by Sheldon and Eleanor Glueck with analysis by Alfred J. Kahn.

The principal recommendation of the book is that an appropriate location for research in social work is as an institute which is part of a graduate professional school of social work in a university. This recommendation is motivated by the criteria of Chapter VI and seems to be both reasonable and feasible. The specific suggestions and financial estimates given in Chapter VII for the establishment and staffing of such an institute for research in social work also appear to be reasonable.

This book represents a well-thought-out first approach to the problem of obtaining valid quantitative information in the important practical field of social work. The basic difficulties are clearly recognized and no attempt is made to underemphasize their magnitude. Slow progress is anticipated in attaining even moderate objectives.

Evidently the development of research practices and theories which have practical value for social work would require a large amount of time, effort, and money. However, the great scope and importance of social work would seem to indicate that the eventual gains from a well-founded evaluative research program would far offset the investment involved. This has been the case in many other practical fields and there is no indication that the social work field represents an exception.

RANDOM DIGITS (15,126–17,375)

From *A Million Random Digits*, to be published by the Rand Corporation, Santa Monica, California.

33591	59785	12833	98932	68064
58418	90331	55858	04015	21454
64446	51017	22280	75597	50227
72136	00303	38880	93327	49522
98626	82484	54610	07211	78610
58393	20225	05436	46172	88951
37346	51007	38032	36002	21080
70712	44236	96795	92351	92844
93585	09918	30983	44282	66849
09473	72923	16747	49802	50639
40229	34921	60405	06803	19332
39795	77221	10012	40798	33864
10288	57483	10881	58984	45136
75702	69428	34047	76224	45887
18129	93659	58389	19715	66259
17777	41004	47057	30688	07539
75195	62294	03371	11672	13089
09722	67635	12114	63055	09214
78800	86912	42076	50287	97998
94287	54751	36242	36557	85664
35997	30761	97081	09501	68887
88241	30402	12318	52430	40139
54382	73370	26184	14024	57444
77681	74946	02099	69091	19372
53148	26074	52293	65359	63971
27212	89889	46933	13364	33883
03867	03105	87912	29610	75108
79895	82633	19209	21548	35022
55256	69386	57453	70147	73538
75937	31113	07607	48037	71020
83389	80236	65972	74528	40888
37363	30345	79933	71058	34826
21960	95585	40374	13239	56162
49562	44137	46625	20031	08524
22666	27414	30980	74485	26480
55386	95918	92481	49234	62616
97725	69513	36950	63526	93824
04703	95851	90956	64424	95979
20077	65817	99523	73180	59978
21343	30031	18840	99260	21284

54021	29008	83672	53679	96395
71820	11033	20183	08804	29493
17709	94849	31771	23244	81585
08254	78963	95437	30231	59108
81351	70961	10255	60437	84576
30665	43699	03593	29165	59990
09069	78653	90094	42735	82876
50593	14698	04737	72551	36417
49578	18100	59836	73221	91696
58388	40915	94507	32209	11548
36995	36202	07971	67001	62062
66348	87666	78055	44485	82955
45457	78252	98239	40000	75563
35406	59553	57852	07506	00009
09678	24538	52426	84852	83781
31604	55850	25644	44972	62275
05523	34355	75127	69797	71419
09458	29207	43632	32905	38513
74086	24077	21369	93541	75329
89299	30765	00348	01134	71581
78089	36887	65867	08595	47390
62067	41248	78667	95282	05622
18924	44409	45345	04972	52794
78656	18550	37171	13483	56058
68420	16483	26482	85964	71336
39182	51174	98146	42953	06606
26224	19972	13442	69662	76755
15737	00496	57764	52052	12835
51093	62290	38948	47473	16618
67799	28342	14551	44149	29854
23875	56766	01932	90063	31883
13701	95168	13169	42055	75609
37741	86434	22400	92732	99851
42479	47405	14055	75427	34464
76691	33263	62048	70917	97697
86707	18895	81790	71294	74832
92564	39987	02283	89970	28776
94648	05598	32171	28793	36746
29838	10664	28050	60122	13409
36368	17792	84792	76594	74110
09254	07510	51039	91683	84500
41486	58524	54508	20707	58504
90794	51934	03295	26582	16300
58558	84833	17105	46659	25003
52392	53546	70291	98846	67315
46095	19501	56181	85351	05023
09925	96974	34321	05454	12862
70550	42483	71204	99628	40642
50281	83708	27298	92651	95086
22767	26514	56322	78635	63731

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 267 SEPTEMBER 1954 *Volume 49*

A QUARTERLY MODEL FOR THE UNITED STATES ECONOMY*

Harold Barger
Columbia University
and
Lawrence R. Klein
University of Michigan

Many theories about economic behavior imply a belief that it can be represented by some system of equations whose solution is determinate. The econometric problem is to specify the form of the functions and to estimate the parameters. It may be that there exist more than one system adequate for the purpose; it may also be that, for an entire economy, even the simplest such system is too complex to be estimated from any known body of data. This paper describes an attempt to represent quarterly movements of gross national product in the U. S. economy by a model with as few as three equations. Failure in the attempt may suggest that a more complex model is required.

Most empirical econometric studies have been based on annual time series data,[1] the sample size seldom exceeding twenty to thirty observations. It has frequently been suggested that econometric research would benefit from the use of quarterly data. The sample would be enlarged by a factor of four and more detailed information obtained about movements of the economy. Although we do not expect to get four times as much information by shifting from annual to quarterly data, we do expect to add something to our knowledge of underlying economic structure. For instance lags can be measured more accurately.

* This paper describes some results of an investigation made possible by grants from the Columbia University Council for Research in the Social Sciences and the Social Science Research Council, to both of whom we are grateful. We are also indebted to Sylvia Schlachter, formerly of Columbia University, and now of the University of Michigan, who carried out the computations.

[1] See, however, Colin Clark, "A System of Equations Explaining the United States Trade Cycle, 1921 to 1941," *Econometrica*, 17 (1949), pp. 93–124, where use is made of quarterly data.

On the other hand we face new problems such as seasonal variation and increased serial correlation of disturbances.

In this paper we shall discuss two simple quarterly models and estimate their parameters from U. S. data for the interwar period. We shall then test the models by extrapolating the results into the period since World War II.

CHOICE OF A MODEL

Our starting point is the three-equation model already fitted by Klein to annual data.[2] This choice was made so that we would have a a good basis for comparison of our results with those of an annual model. The three-equation model is compact and permits considerable experimentation. Of the following variables, all of which are measured in constant prices, the first six are regarded as endogenous and the remaining as exogenous:

C	consumer expenditures
W_1	wages and salaries paid by private industry
Π	non-wage income or "profits"
I	net private domestic investment
K	year-end stock of capital
Y	net national income
W_2	wages and salaries paid by government
T	indirect taxes less subsidies
G	government purchases plus net foreign balance
t	time
u	random disturbance

In the annual model, equations for which are as follows, negative subscripts outside parentheses denote variables lagged by the number of *years* indicated.

$$C = \alpha_0 + \alpha_1(W_1 + W_2) + \alpha_2\Pi + \alpha_3(\Pi)_{-1} + u_1 \tag{1.1}$$

$$I = \beta_0 + \beta_1\Pi + \beta_2(\Pi)_{-1} + \beta_3(K)_{-1} + u_2 \tag{1.2}$$

$$W_1 = \gamma_0 + \gamma_1(Y + T - W_2) + \gamma_2(Y + T - W_2)_{-1} + \gamma_3 t + u_3 \tag{1.3}$$

$$Y + T = C + I + G \tag{1.4}$$

$$Y = \Pi + W_1 + W_2 \tag{1.5}$$

$$I = K - (K)_{-1} \tag{1.6}$$

[2] Lawrence R. Klein, *Economic Fluctuations in the United States, 1921–1941* (New York: Wiley, 1950), pp. 55–80.

Of the three stochastic equations, (1.3) can be regarded either as distributing income between wages and profits (including interest and rent), or as the demand for labor. If we prefer the latter viewpoint, a direct measure of private output is required; therefore we replace $(Y+T-W_2)$ by $(C+I+G-W_2)$.[3] The last three equations, being merely accounting identities, have known coefficients and are not subject to random disturbances in behavior. Equation (1.4), however, is subject to errors of observation insofar as a statistical discrepancy exists between direct estimates of national product (expenditure version) and estimates of national income (factor payments version) converted to market prices.

To adapt the above model so that the variables relate to quarters instead of years, a one-year lag needs to be replaced by a lag distributed over several quarters. Yet if lagged variables are introduced too freely into a linear scheme, intercorrelation between them may render results indeterminate. Therefore we grouped lagged variables in pairs (from this point forward, negative subscripts outside parentheses will denote values lagged by the number of *quarters* indicated):

$$(x)_{-(2n+1)/2} \equiv \frac{(x)_{-n} + (x)_{-(n+1)}}{2}\,.$$

With the help of this convention we might (for instance) write the following quarterly version of (1.1) to (1.6), the time unit referred to being three months instead of a year.

$$(2.1)\qquad C = \alpha_0 + \alpha_1(W_1 + W_2) + \alpha_2\Pi + \alpha_3(\Pi)_{-3/2} + u_1,$$

$$(2.2)\qquad I = \beta_0+\beta_1\Pi+\beta_2(\Pi)_{-3/2}+\beta_3(\Pi)_{-7/2}+\beta_4(\Pi)_{-11/2}+\beta_5(K)_{-1}+u_2,$$

$$(2.3)\qquad W_1 = \gamma_0+\gamma_1(C+I+G-W_2)+\gamma_2(C+I+G-W_2)_{-3/2}+\gamma_3 t+u_3,$$

together with (1.4), (1.5), and (1.6) above.

THE ESTIMATION OF PARAMETERS

The equations (2.1), (2.2), and (2.3) each contain at least two endogenous variables. We could, of course, estimate the coefficients in each equation by conventional use of least squares. To do so we should have arbitrarily to treat some particular variable as dependent, and all others (endogenous or otherwise) as independent. This procedure is arbitrary and, moreover, is known to lead to biased estimates of the

[3] Were it not for the statistical discrepancy in our national accounts, we would have the same measure of output reckoned as a sum of expenditures or as a sum of factor income payments, and the replacement mentioned would not affect computations.

parameters.[4] To be sure, in at least one case that has been investigated, i.e. the annual model given above, the bias proved to be smaller for most parameters than the estimated sampling errors.[5] We do not know that this will always be the case. In any event the compulsion to choose as dependent one only among several variables, all of which clearly are endogenous, is unwelcome. Therefore we decided to use unbiased or consistent methods of estimation.

To obtain consistent estimates of the parameters in (2.1) to (2.3), the equations must be regarded as simultaneous and the system treated as a whole. With six endogenous and nine exogenous or predetermined variables, estimation by the limited-information maximum-likelihood method is fairly laborious; yet this system is the closest analogue to the annual model, and computation of its parameters was undertaken.[6] The results contradicted the assumption that the disturbances (u_1, u_2 and u_3) were random. Quarterly are more highly autocorrelated than annual variables, and it is not surprising that the same should be true of the residuals from a regression between such variables.

An obvious device is to assume that the disturbances satisfy a lag correlation scheme such as

$$u_i = \sum_{j=1}^{3} \rho_{ij}(u_j)_{-1} + v_i \qquad (v_i \text{ mutually independent}).$$

However, to estimate ρ_{ij} simultaneously with the other parameters by the method of maximum likelihood would be burdensome.

In the limited information method, each individual equation of the system uses only restrictions on the parameters of that equation and not those on other equations. Thus it would be possible to obtain limited information estimates if each disturbance satisfied a pure autoregressive equation

$$u_i = \rho_i(u_i)_{-1} + v_i \qquad (v_i \text{ mutually independent}).$$

Yet the computational burden imposed even with this simplification was more than we wanted to undertake at this stage of research.

A RECURSIVE MODEL

Our procedure was somewhat different. We converted the system of equations to recursive form, thus enabling us to obtain consistent

[4] See, e.g., T. Koopmans, "Statistical Estimation of Simultaneous Economic Relations," *Journal of the American Statistical Association*, 40 (1945), pp. 448–66.

[5] Klein, *op. cit.* Compare pp. 68, 75.

[6] For computational procedure see, e.g., T. W. Anderson and H. Rubin, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, XX (1949), pp. 46–63.

estimates by repeated application of single-equation least-squares methods.[7] It proved computationally feasible to do this while assuming that the disturbances satisfied first-order autoregressive equations. For this arrangement the first equation to be estimated must have only one endogenous variable, the remaining variables being predetermined or exogenous. Each subsequent equation must contribute only one additional endogenous variable. In estimating an equation that contains more than one endogenous variable, in the case of all but one such variable the calculated rather than the observed values are used in the computations. The procedure leads to consistent estimates. In some recursive systems, the assumption that unlagged disturbances in separate equations are independent makes maximum likelihood estimates obtainable by repeated application of the method of least squares. If we drop these independence assumptions and substitute calculated instead of observed values of endogenous variables in successive equations, we obtain consistent but not necessarily full maximum-likelihood estimates.

From a formal standpoint it makes no difference which equation we estimate first, provided it can be written with one endogenous variable as a function of predetermined variables alone. We believe that investment decisions depend upon a longer range of past experience, and result more slowly in actual expenditures, than other types of decision. Therefore we put $\beta_1 \equiv$ zero in equation (2.2), estimate the remaining parameters in this equation by least squares, and use the calculated value of investment as an exogenous variable in the consumption equation. This treatment allows consumption to respond immediately to changes in income and permits consistent estimation by least squares. On the other hand, the distinction between wage and nonwage income, as in equation (2.1), has to be abandoned. On substituting calculated (denoted by superior $\widehat{\ }$) for observed values of I,

$$\widehat{Y} = C + \widehat{I} + G - T;$$

and we may write:

$$\begin{aligned} C &= \alpha_0' + \alpha_1'\widehat{Y} + u_1' \\ &= \frac{\alpha_0'}{1 - \alpha_1'} + \frac{\alpha_1'}{1 - \alpha_1'}(\widehat{I} + G - T) + u_1'' \\ &= \alpha_0'' + \alpha_1''(\widehat{I} + G - T) + u_1''. \end{aligned}$$

The disadvantage resulting from the consolidation of wage and nonwage income may be partially offset by using Y', disposable income,

[7] Herman Wold and L. Juréen, *Demand Analysis* (New York: Wiley, 1953), p. 14.

instead of Y, national income.[8] We shall further introduce lagged consumption on the right hand side, to take account of the influence of past behavior.[9] Finally we substitute computed values on the right hand side of the wage equation (2.3) and estimate this also by the method of least squares.

The above procedures yield the following fully recursive model which may be estimated consistently by least squares methods.

Model A (Fully recursive)

$$(3.1)\quad \begin{cases} C = \alpha_0 + \alpha_1(C)_{-3/2} + \alpha_2(\hat{I} + G - T') + u_1 \\ u_1 = \rho_1(u_1)_{-1} + v_1 \end{cases}$$

$$(3.2)\quad \begin{cases} I = \beta_0 + \beta_1(\Pi)_{-3/2} + \beta_2(\Pi)_{-7/2} + \beta_3(\Pi)_{-11/2} + \beta_4(K)_{-1} + u_2 \\ u_2 = \rho_2(u_2)_{-1} + v_2 \end{cases}$$

$$(3.3)\quad \begin{cases} W_1 = \gamma_0 + \gamma_1(\hat{C} + \hat{I} + G - W_2) + \gamma_2(C + I + G - W_2)_{-3/2} + \gamma_3 t + u_3 \\ u_3 = \rho_3(u_3)_{-1} + v_3 \end{cases}$$

$$(3.4)\quad Y' + T' = Y + T = C + I + G$$

$$(3.5)\quad Y' + T' - T = \Pi + W_1 + W_2$$

$$(3.6)\quad I = K - (K)_{-1}$$

Equation (3.2) must be estimated first, and the calculated values for $\hat{I}$ substituted in (3.1). When (3.1) has been estimated, values of $\hat{C}$ and $\hat{I}$ are substituted in (3.3). The complication introduced by the autoregressive treatment of the disturbances is discussed in the Appendix.

More generally, it may be seen that the condition for recursive treatment (i.e. consistent estimation in stepwise fashion) is that the Jacobian of the transformation connecting the disturbances with the endogenous variables shall be triangular and (by the rule of normalization) equal to a constant, unity. Thus, in the present example

$$\frac{\partial(v_2, v_1, v_3)}{\partial(I, C, W_1)} = \begin{vmatrix} 1 & -\alpha_2 & -\gamma_1 \\ 0 & 1 & -\gamma_1 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

The reader will observe that the first column of the Jacobian refers to

[8] Y' is obtained from Y by adding net interest paid by government and government transfer payments; and deducting personal tax payments, corporate tax accruals, and all contributions for social insurance (concepts follow Department of Commerce practice). In equations (1.1) and (2.1), data did not allow the wage and nonwage income separately to be placed upon a disposable basis.

[9] This type of lag relation has been found to be satisfactory in models based on annual data. See T. M. Brown, "Habit Persistence and Lags in Consumer Behavior," *Econometrica*, 20 (1952), pp. 355–72.

equation (3.2), estimated first; the second column to equation (3.1), estimated next; and the third column to equation (3.3), estimated last.

A HYBRID MODEL

We saw above that the simultaneous estimation of all three equations requires the assumption that disturbances are random, and that in practice with quarterly data this assumption is contradicted. To avoid this difficulty we developed a recursive model, as just explained. A further alternative is to use a model which is partly recursive and partly simultaneous. Thus we may estimate $\hat{I}$ as above and substitute the calculated values in the remaining two equations, and then proceed to estimate the latter simultaneously by the method of limited information (instead of successively by the method of least squares). If we take our consumption equation

$$C = \alpha_0' + \alpha_1'(C)_{-3/2} + \alpha_2'(W_1 + W_2) + \alpha_3'\Pi + u_1',$$

and for Π write

$$\hat{\Pi} = C + \hat{I} + G - W_1 - W_2 - T,$$

we obtain (4.1) below.

Model B (Hybrid)

$$(4.1) \qquad C = \alpha_0 + \alpha_1(C)_{-3/2} + \alpha_2(W_1 + W_2) + \alpha_3(\hat{I} + G - T) + u_1$$

$$(4.2) \qquad \begin{cases} I = \beta_0 + \beta_1(\Pi)_{-3/2} + \beta_2(\Pi)_{-7/2} + \beta_3(\Pi)_{-11/2} + \beta_4(K)_{-1} + u_2 \\ u_2 = \rho_2(u_2)_{-1} + v_2 \end{cases}$$

$$(4.3) \qquad W_1 = \gamma_0 + \gamma_1(C + \hat{I} + G - W_2) + \gamma_2(C + I + G - W_2)_{-3/2} + \gamma_3 t + u_3$$

$$(4.4) \qquad Y + T = C + I + G$$

$$(4.5) \qquad Y = \Pi + W_1 + W_2$$

$$(4.6) \qquad I = K - (K)_{-1}$$

Here consumption depends separately upon wage and nonwage income, as in the original annual model, and this feature makes fully recursive treatment impossible. Equation (4.2) is estimated by the method of least squares as before, but even after calculated values for I have been substituted, equations (4.1 and (4.3) both contain more than one endogenous variable. However we estimated (4.1) and (4.3) consistently by doing so simultaneously, using computational procedures to which reference has already been made.[10]

[10] Anderson and Rubin, *op. cit.* Equations (4.1) and (4.3) do not contain autoregressive error terms because of the heavy additional burden of computation that their introduction would impose: see discussion above.

SEASONAL VARIATION IN QUARTERLY DATA

In the later sections of this paper the parameters in Models A and B are estimated, and their predictive power tested, using seasonally adjusted data throughout. The practical necessity of using corrected data can readily be demonstrated. In the first place, some components of national income (e.g. farm operators' income) already have been partially corrected, and other components (e.g. business profits) have been fully corrected for seasonal variation during original collection or subsequent processing of the data; such series are not available in "seasonally unadjusted" form. The implications of using data, some of which have and some of which have not been corrected for seasonal movement, are obscure.

In the second place, if seasonally unadjusted data are to be used, the seasonal behavior has to be incorporated into the model itself. A simple and plausible representation of seasonal behavior requires multiplication by a parameter that varies only with the season. Thus if

$$Y = \alpha_1 z_1 + \alpha_2 z_2 + \cdots + u \tag{5.1}$$

is a structural equation in the absence of seasonal variation, to allow for such variation in the model we need to write

$$Y = (\beta_1 z_1' + \beta_2 z_2' + \beta_3 z_3' + \beta_4 z_4')(\alpha_1 z_1 + \alpha_2 z_2 + \cdots) + u \tag{5.2}$$

where z_i' assumes the value unity in the ith quarter and is zero in all other quarters. If the β as well as the α have to be estimated from the data, it will be found that the estimating equations do not readily admit of numerical solution since the behavior equation (5.2) is nonlinear in the parameters.[11] An advantage of using unadjusted data in this way would be knowledge of the number of degrees of freedom absorbed in estimating the β. The adjustment of data for seasonal variation also uses up degrees of freedom, but we never know just how many degrees are absorbed in the process.

However, theoretical superiority does not seem to lie wholly on the side of unadjusted data. It may be urged that the economic subject makes his own (rough) seasonal corrections as he goes along. The consumer does not react to his income and to the time of year as separate data, but asks himself, "Is this more or less than the income I would expect at this time of year?" The entrepreneur looks at his profits and

[11] See also L. Hurwicz, "Variable Parameters in Stochastic Processes: Trend and Seasonality," *Statistical Inference in Dynamic Economic Models*, ed. T. C. Koopmans (New York: Wiley, 1950), Ch. XI.

compares them with some level expected for the season. Each applies a rough seasonal correction he carries in his head.

If this version of the facts is accepted, seasonally adjusted rather than unadjusted data appear to be the more accurate measure of the variables that interest us. However, the seasonal correction effected by the economic subject must be based wholly on past experience, although it doubtless is continually revised as history unfolds. Of course, seasonal corrections made by statisticians are based on data for the current year, earlier years, *and later* years. The theoretical justification for using standard methods to correct the raw data for seasonal variation is therefore incomplete. The most we can say is that seasonal adjustment by conventional methods enables us to approximate more closely the variables most relevant to economic behavior than no seasonal adjustment at all.

SAMPLE DATA USED FOR ESTIMATION

Parameters in both models A and B were estimated from 72 quarterly observations for the years 1923–1940. All observations were expressed in constant (1939) prices. Variables were defined in accordance with Department of Commerce usage[12] except that we substituted a fresh series for capital consumption allowances. Annual Commerce figures for 1929–1938 were interpolated with data from Barger's *Outlay and Income in the United States*[13] and were extrapolated back to 1921 in the same way. All data are seasonally adjusted quarterly totals and are expressed in \$ million in 1939 prices,[14] excepting only t which numbers the quarters consecutively.[15]

MODEL A (FULLY RECURSIVE)

Investment equation. Equation (3.2) has the following coefficients when estimated by least squares.

---

[12] See *Survey of Current Business*, "National Income Supplements." Π comprises business profits (corporate and noncorporate, including income of farmers and the independent professions), and interest and rents, before tax.

[13] New York: National Bureau of Economic Research, 1942.

[14] The various components of gross national product were deflated with readily available price indexes. Capital consumption allowances were deducted from gross national product (both in 1939 prices), yielding net national product. A comparison of the latter, quarter by quarter with net national product in current prices yielded a single implicit price index which was used to deflate all components of income.

[15] To print the sample data would require excessive space. Magnitudes may be indicated by quoting the following mean values for the sample period: C, 14,538; W_1, 9,784; Π, 4,126; I, 306; Y, 15,253; W_2, 1,343; T, 1,674; G, 2,396; Y', 16,173. Values of K and t are zero for the last quarter of 1922; and 22,030 and 72 respectively for the last quarter of 1940.

$$(6.1)\quad \begin{cases} I = -2016 + \underset{(0.403)}{0.438}(\Pi)_{-3/2} + \underset{(0.360)}{0.615}(\Pi)_{-7/2} - \underset{(0.429)}{0.205}(\Pi)_{-11/2} \\ \qquad - \underset{(0.038)}{0.054}(K)_{-1} + u_2 \\ u_2 = \underset{(0.211)}{0.603}(u_2)_{-1} + v_2 \end{cases}$$

In parentheses are shown estimates of the sampling errors, whose exact distributions are not known. While sampling errors for individual coefficients of $(\Pi)_{-3/2}$, $(\Pi)_{-7/2}$, and $(\Pi)_{-11/2}$ are quite large, the corresponding error for their sum is smaller than any of the separate errors, being 0.294. The statistic δ^2/s^2 $(=2.016)$ may be computed as the ratio of the mean square successive difference of the residuals v_2 to their variance. The distribution of the statistic is such that for a random series with (say) 60 degrees of freedom the probability is 0.95 that δ^2/s^2 will exceed 1.6.[16] Hence the result is compatible with the assumption that the disturbances are random.

Consumption equation. As the next step we substitute $\hat{I}$ from (6.1) on the right hand side of (3.1). Estimation of the latter equation yielded results compatible with the assumption of random v_1, but reported a negative value for α_2. This result implies that consumption depends inversely on current income, a conclusion we rejected. Deciding that α_2 must be positive, we put ρ_1 identically equal to zero, and contented ourselves with the estimate

$$(6.2)\qquad C = 266 + 0.990(C)_{-3/2} + 0.036(\hat{I} + G - T') + u_1$$

which is equivalent to

$$(6.3)\qquad C = 257 + 0.955(C)_{-3/2} + 0.035\hat{Y}' + u_1'$$

for which $\delta_2/s_2 = 1.35$, indicating significant autocorrelation of the residuals. An autoregression coefficient computed from the observed residuals of (6.2) was $\hat{\rho}_1 = 0.329$.

Because of the transformation of variables, the confidence intervals for the coefficients in (6.3) are not symmetrical about the point estimates. If α_1' is the coefficient of $(C)_{-3/2}$ and α_2' of $\hat{Y}'$ in (6.3), the limits at the 95% level are approximately

[16] B. I. Hart and J. von Neumann, "Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Variance," *Annals of Mathematical Statistics*, XIII (1942), pp. 207–14. We of course use only the lower tail of the distribution.

$$0.931 \leqq \alpha_1' \leqq 0.984; \qquad \text{and} \qquad -0.072 \leqq \alpha_2' \leqq 0.122.$$

For Model A this consumption function was the least implausible we obtained in a number of trials, with and without lagged values of consumption, with and without autoregressive transformation of disturbances. Plainly it suffers from three serious defects. (1) The residuals cannot be considered random. (2) The low value of α_2 (or α_2') furnishes but a very weak link with the investment equation. (3) The confidence limits for α_2 (or α_2') include negative values.

The marginal propensity to consume. Despite the fact that $(C)_{-3/2}$ is the dominating variable in (6.3), Y' does play a larger role as the time over which the equation functions is lengthened. Let

$$(C)_1 = \alpha_0 + \alpha_1(C)_0 + \alpha_2(Y')_1 + (u)_1.$$

Then

$$(C)_n = \alpha_0 \sum_{i=0}^{n-1} \alpha_1{}^i + \alpha_1{}^n(C)_0 + \alpha_2 \sum_{i=0}^{n-1} \alpha_1{}^i(Y')_{n-i} + \sum_{i=0}^{n-1} \alpha_1{}^i(u)_{n-i}.$$

In the limit, if we assume a steady income stream, we have

$$C = \frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_2}{1-\alpha_1} Y' + v \tag{6.4}$$

where v is a linear combination of random variables. Evidently we may regard $\alpha_2/(1-\alpha_1)$ as a long-run marginal propensity to consume, its computed value being about 0.78. We therefore report a sizeable influence of income upon consumption in the long run, even though (6.3) has the appearance of almost pure autoregression.

We can also construct a joint confidence region for α_1 and α_2 in the form of an ellipse, to see whether $\alpha_2/(1-\alpha_1)$ is estimated in the range (0, 1) even though α_2 is not itself significantly positive.[17] Our estimate of the propensity to consume proves very rough, for about one-tenth of the area of an ellipse drawn at the 95% level admits negative values.

The wage equation. The calculation of (3.3) yielded an estimate of ρ_3 practically equal to unity and substantial autocorrelation of the residuals v_3. Moreover $(\gamma_1+\gamma_2)$ was estimated well below 0.4, although annual models have shown the marginal influence of $(C+I+G-W_2)$ to be between 0.5 and 0.6. Neither estimation of the equation in terms of first differences (i.e. putting ρ_3 identically equal to unity), the addi-

[17] An example of the preparation of this type of confidence region is given in T. Haavelmo, "Methods of Measuring the Marginal Propensity to Consume," *Journal of the American Statistical Association*, 42 (1947), pp. 105–22.

tion of longer lags for the independent variable, the insertion of higher powers of t, nor the use of a second-order autoregressive scheme for the disturbances, led to improved results. Nor were lagged values of the dependent variable helpful on the right hand side of the equation. Probably the distribution of income cannot be satisfactorily approximated by any simple linear relation. In the present study we contented ourselves with the following equation in which ρ_3 is identically zero:

$$\text{(6.5)}\quad W_1 = \underset{}{665} + \underset{(0.210)}{0.759}(\hat{C}+\hat{I}+G-W_2) - \underset{(0.212)}{0.178}(C+I+G-W_2)_{-3/2} - \underset{(4.3)}{0.953}t + u_3$$

The estimated coefficients are reasonable enough, but $\delta^2/s^2 = 1.21$ (i.e. clearly <1.6). The residuals therefore are autocorrelated, and in fact yield an autoregression coefficient of 0.401. Despite a common belief that the distribution of income was shifting during the period that we chose for our sample, the coefficient of t proves not to be significantly different from zero.

MODEL B (HYBRID)

Investment equation. It will be recalled that equations (4.2) and (3.2) are identical, investment being estimated in the same fashion in both models. Calculated values, $\hat{I}$, are therefore obtained from (6.1) above.

Consumption and wage equations. Parameters in equations (4.1) and (4.3) need to be estimated by the simultaneous treatment of these two equations.[18] The results, shown as (6.6), (6.7), and (6.8), are consistent but not efficient. Because endogenous variables appear on the right hand side, estimation of (4.1) and (4.3) independently by least squares leads to biased results. Estimates of the coefficients obtained separately for each equation by least squares are shown in (6.7) and (6.8) in parentheses *above* the unbiased estimates, and suggest that the bias is not quantitatively important in the present case. Sampling errors of the unbiased estimates are shown in parentheses *below* the latter.

$$\text{(6.6)}\quad C = 1061 + \underset{(0.174)}{0.721}(W_1+W_2) + \underset{(0.146)}{0.383}(C)_{-3/2} - \underset{(0.065)}{0.192}(\hat{I}+G-T) + u_1$$

which is equivalent to

$$\text{(6.7)}\quad C = \overset{(1375)}{1313} + \overset{(+0.696)}{0.655}(W_1+W_2) + \overset{(+0.443)}{0.474}(C)_{-3/2} - \overset{(-0.257)}{0.238}\hat{\Pi} + u_1'$$

[18] For computational procedure, see Anderson and Rubin, *op. cit.*

$$\begin{array}{lll} & (570)(+0.546) & (-0.041) \\ (6.8)\ W_1 = 604 & +0.706(C+\hat{I}+G-W_2) & -0.117(C+I+G-W_2)_{-3/2} \\ & (0.098) & (0.096) \\ & (-1.67) & \\ & -2.815t+u_8 & \\ & (1.97) & \end{array}$$

As before we calculated the residuals and estimated ρ from them.

For (6.6)

$$\frac{\delta^2}{s^2} = 1.67 \qquad \widehat{\rho_1} = 0.177,$$

and for (6.8)

$$\frac{\delta^2}{s^2} = 1.45 \qquad \widehat{\rho_3} = 0.288.$$

For the consumption equation the result is just compatible with the assumption of random disturbances, but the wage equation, as in Model A, yields residuals that show significant autocorrelation. As in Model A, the investment equation is not solidly linked to the others: although lagged consumption plays a smaller role than in (6.2) and (6.3), the coefficients of $\hat{I}$ in (6.6) and of $\widehat{\Pi}$ in (6.7) have the wrong sign.

TESTING THE MODELS

How well do our models represent behavior, first during and second outside the sample period? The acid test is the model's ability to predict. Of course the forecasts of models such as ours are *conditional*, in the sense that the values of the exogenous (but not of the endogenous) variables must be known for the period to which the forecast applies. To be sure the practical forecaster may regard this as a fatal disadvantage; here, however, we are concerned with prediction, not for its own sake, but only as a means of testing the models. The sample period comes to an end in 1940. Accordingly we shall test the model by making predictions for 1941 and for 1947 through 1952, omitting the war years as not relevant for the purpose in hand.

The procedure is as follows. We can estimate the values of the endogenous variables for any quarter (say, 1st of 1947) on the basis of observed values of all quantities for preceding quarters (through 4th of 1946) and data for exogenous variables only (W_2, T or T', and G) for

the current quarter (1st of 1947), and compare the estimates so obtained with the observed values of the endogenous variables for the current quarter (1st of 1947). We can repeat this for the 2nd quarter of 1947, using complete data for the first quarter and exogenous variables for the second quarter, and so on. Thus we build up a succession of short-range predictions, each one quarter ahead. The question may then be posed, whether the performance of the model is appreciably better than guesswork.

Solution of the three equations for each model, together with the accounting identities, yields estimates of six quantities: gross national product (GNP),[19] national income, and the two endogenous components of each (I, C; W_1, Π). Of course the six predictions are not independent, for the accounting identities allow any three predicted variables to be derived from the remaining three. Because the investment equation stands by itself, in that it predicts a single endogenous variable in terms of predetermined variables, and because of the obvious interest attaching to predictions of GNP, we shall for the sake of brevity report the quarter-by-quarter tests only for investment and GNP.[20] The outcome of the long-range extrapolation to be mentioned later will be summarized for all six variables.

In the following tabulations, s is the root mean square difference between predicted and observed values; p the proportion of cases in which the direction of change from the preceding quarter (i.e. the sign of the first difference) is correctly predicted; and P the probability of so high a proportion of directions of change being correctly predicted by chance, on the assumption that correct and incorrect predictions are equally likely, and that probabilities of correct prediction in successive quarters are independent. Clearly where variables are known to be autocorrelated, guesswork can easily improve upon the toss of a coin, and probabilities of correct prediction in successive quarters no longer are independent. Therefore we have computed p and P, not only for all quarters taken together, but also for those particular quarters in which the variable to be predicted was later observed to have reversed its direction. For this more stringent test, the chances of correct prediction in individual quarters appear to be completely independent. In simple language, it perhaps is not too hard to guess that an upward or down-

[19] To predict GNP, we also need to know depreciation, assumed to be predetermined.

[20] Results of quarter-by-quarter tests for consumption and national income were similar to those about to be given for investment and GNP—as indeed they must be in view of the accounting identities. Short-range predictions for wage and nonwage income were inferior and could be more readily explained by chance, a result perhaps connected with the shift in the distribution of income since World War II referred to below.

ward sweep will continue; it is less easy correctly to predict reversals of direction.

Predicting investment. The investment equation (6.1), common to both models A and B, yields the results in Table 1. In predicting 28 of the 40 reversals of direction during the sample period, the investment equation obviously does much better than could have been achieved by accident. The result during 1947–1953 is less impressive, for only 7 of 9 reversals were correctly predicted, the chance of so good a result occurring by accident being about 1 in 10.

TABLE 1

INVESTMENT: SUCCESSIVE QUARTERLY PREDICTIONS
(s in $ million at 1939 prices)

All Quarters			Quarters in Which Direction of Change was Reversed	
s	p	P	p	P
		1923 to 1940 (sample period)		
477	51/72	.001	28/40	.008
		1941		
343	3/4	.31	0/1	1.00
		1947 to 1952		
567	17/24	.032	7/9	.09

Since the investment equation (3.2) can be written in the form

$$\begin{aligned} I = \rho_2(I)_{-1} &+ \beta_0(1-\rho_2) + \beta_1[(\Pi)_{-3/2} - \rho_2(\Pi)_{-5/2}] \\ &+ \beta_2[(\Pi)_{-7/2} - \rho_2(\Pi)_{-9/2}] + \beta_3[(\Pi)_{-11/2} - \rho_2(\Pi)_{-13/2}] \\ &+ \beta_4[(K)_{-1} - \rho_2(K)_{-2}] + v_2, \end{aligned}$$

we can estimate the variance of forecast values of I as a sum of products, each product being obtained by multiplying an estimated variance or covariance of β_1, β_2, β_3, β_4, and ρ_2 (or of their products) by a square or product of values of predetermined variables in the forecast period, together with an estimate of the residual variance.[21] Taking as values of the predetermined variables their mean level during the 24 forecast quarters 1947–1952, we can approximate a standard error of forecast

[21] See, e.g., H. Hotelling, "Problems of Prediction," *American Journal of Sociology*, 48 (1942–43), pp. 61–76.

for the postwar period of $907 million (1939 prices). The root mean square of the observed errors of prediction for 1947–1952 (s in Table 1) was somewhat smaller than this.

Predicting gross national product. The performance of the complete models A and B is conveniently tested by predicting GNP.[22] We did not estimate the standard error of forecast for GNP because of the heavy computation required. But we carried out the other tests described above, both for models A and B and for three "guesswork models" of progressively increasing sophistication. Call GNP y and indicate lags, as before, by suffixes. Assume first that GNP will be the same this quarter as last quarter:

(*Guesswork Model I*) $$y = (y)_{-1}.$$

Embodying the simplest possible assumption, this model cannot of course be used for forecasting direction of change, but it affords a criterion for measuring errors of prediction.

Assume, second, that GNP's change from last quarter will be one-half GNP's change last quarter from the preceding quarter:

(*Guesswork Model II*) $$y = (y)_{-1} + \tfrac{1}{2}[(y)_{-1} - (y)_{-2}].$$

The fraction one-half is chosen arbitrarily on the assumption that for a stable series such as GNP the autoregression coefficient between first differences lies between 0 and 1.[23]

Thirdly, let us fit a second-order difference equation to the observed values of GNP during the sample period (1923–1940):

(*Guesswork Model III*) $$y = 4828 + 0.873(y)_{-1} - 0.125(y)_{-2}.$$

Model III is perhaps no longer pure guesswork, but it still will serve as a standard of performance against which to test our two econometric models A and B. Results of the tests are shown in Tables 2, 3, and 4.

During the sample period, both of the econometric models fit the observed data markedly better than any of those based on guesswork (Table 2). However, in predicting reversals of direction, only Model A performs significantly better than might be expected from chance.

---

[22] GNP equals $(C+I+G)$ together with depreciation, all in 1939 prices. C and I are predicted by the model; G and depreciation are exogenous. When predicting with models A and B, we use the respective values of ρ_1 and ρ_2 obtained from residuals observed during the sample period and quoted earlier.

[23] Carl Christ and Milton Friedman have each endorsed a coefficient of unity for this test (National Bureau of Economic Research, *Conference on Business Cycles*, New York, 1951, pp. 57, 110). But unity is an extreme value in the context, and seems unlikely to give as good predictions as some coefficient near the middle of the range (0, 1).

TABLE 2

GROSS NATIONAL PRODUCT: SUCCESSIVE QUARTERLY PREDICTIONS, 1923–1940 (SAMPLE PERIOD)

(s in $ million at 1939 prices)

	All 72 Quarters			32 Quarters in Which Direction of Change Was Reversed	
	s	p	P	p	P
Econometric Model					
A	652	54/72	.0001	24/32	.003
B	708	51/72	.0002	20/32	.108
Guesswork Model					
I	760	36/72	.5	16/32	.57
II	805	40/72	.2	0/32	1.00
III	939	36/72	.5	15/32	.7

During 1941—the year immediately following the close of the sample period—GNP rose steadily, so that no reversals of direction occurred (Table 3). Here, because the period contains no turning point, Guesswork Model II performs as well as the econometric models.

TABLE 3

GROSS NATIONAL PRODUCT: SUCCESSIVE QUARTERLY PREDICTIONS, 1941

(s in $ million at 1939 prices)

	s	p	P
Econometric Model			
A	425	4/4	.06
B	1,857	4/4	.06
Guesswork Model			
I	1,118	2/4	.7
II	594	4/4	.06
III	3,115	0/4	.94

We also carried out the tests for the postwar period, with the results shown in Table 4. Again the econometric models perform much better than any of the attempts at guesswork. Although during 1947–1952

TABLE 4

GROSS NATIONAL PRODUCT: SUCCESSIVE QUARTERLY PREDICTIONS, 1947–1952

(s in $ million at 1939 prices)

	All 24 Quarters			11 Quarters in Which Direction of Change Was Reversed	
	s	p	P	p	P
Econometric Model					
A	651	19/24	.003	8/11	.113
B	1,087	18/24	.011	9/11	.033
Guesswork Model					
I	678	12/24	.58	5½/11	.61
II	687	13/24	.42	0/11	1.0
III	4,499	6/24	.99	5/11	.73

they perform less satisfactorily than during the sample period, both Models A and B show significant predictive power if all 24 predictions are considered independent. On the other hand if the test is confined to the prediction of reversals of direction in GNP, the result is less clear cut. For Model B, chances are 30 to 1 against accidentally calling 9 of the 11 turning points; but for Model A the chances are only 10 to 1 against calling 8 of the turns. Yet the consistently superior performance of Model A, when measured by root mean square error of prediction, is noteworthy.

A LONG-RANGE EXTRAPOLATION

Instead of a series of short-range predictions, each one quarter into the future, we may project a single extrapolation as far ahead as we please, provided only that the exogenous variables are known or can be estimated. Suppose a computer to sharpen his pencil after close of business on December 31, 1946. He is supplied with the model, with data for all variables up to and including the quarter now ended, and with (correctly) anticipated values of the exogenous variables for each quarter through the end of 1952. His forecasts of investment and GNP (both in 1939 prices) are shown in Charts I and II. Predicted investment rises sharply to a peak in 1948 two quarters earlier than the downturn in actual investment. Predicted investment then declines

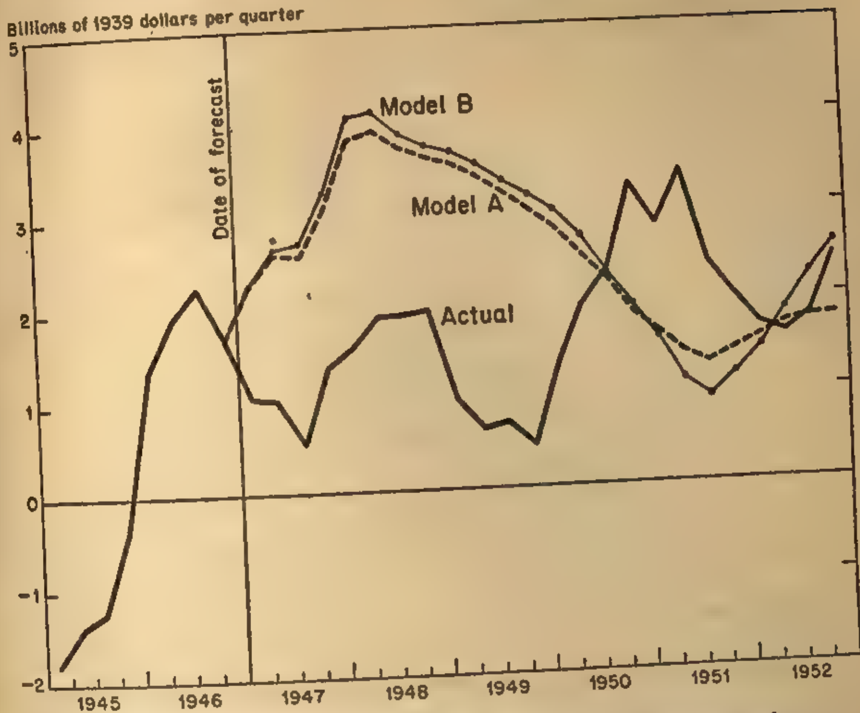


CHART I. Net Private Domestic Investment, as actually observed and as predicted by Models A and B. The predictions are assumed to be made as of Dec. 31, 1946.

steadily for three years, missing entirely the pre- and post-Korea boom, turning up again only in the summer of 1951. In the case of GNP, the upward trend from 1947 to 1952 is correctly predicted, but the wave-like movement forecast for 1949–50 did not eventuate.

Summary totals for each variable—actual and predicted—for the six-year period are shown in Table 5.

For GNP, root mean square differences (s) between prediction and observation for the 24 quarters are: Models A and B, \$2.1 and 2.7 billion respectively; Models I, II, and III, \$3.9, 5.6, and 15.8 billion respectively.

Although the econometric models predict six-year totals for GNP and national income better than any of the guesswork models, the same is not uniformly true of the components. Indeed Guesswork Model I, whose performance elsewhere was so indifferent, happens here to score a bulls-eye in forecasting the six-year total for investment!

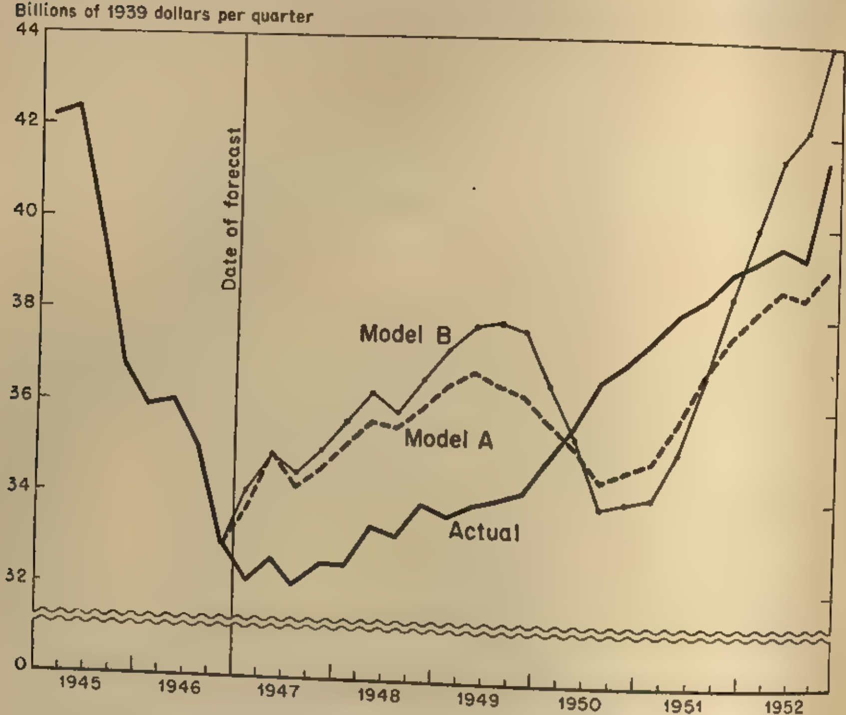


CHART II. Gross National Product, as actually observed and as predicted by Models A and B. The predictions are assumed to be made as of Dec. 31, 1946.

TABLE 5

LONG-RANGE PREDICTIONS, 1947–1952

(Six-year totals in $ billion at 1939 prices)
Assumed date of forecast: December 31, 1946

	I	C	GNP	W_1	Π	Y
As actually observed	41	576	855	451	171	694
Econometric model						
A	60	566	865	421	210	704
B	63	584	885	434	218	724
Guesswork model						
I	40	546	790	435	143	650
II	27	522	740	419	138	622
III	not computed		493	not computed		

Even so unsophisticated a guess must sometimes turn out correct.

Some of the errors of the econometric models can readily be rationalized. Indeed, the diagnosis of errors is very necessary if models are to be improved. In the present instance, the overstatement of investment by the econometric models may be explained by our use of nonwage income before taxes, Π, as independent variable; taxes (especially here corporate taxes) are notoriously higher today than during the sample period.[24] Moreover Π itself is overstated, and this is related to the understatement of W_1; a shift appears to have occurred in the distribution of income since the close of the sample period. To adapt the models to take account of these particular changes would not be difficult, but such revisions lie outside the scope of this paper.

COMPARISON OF ANNUAL WITH QUARTERLY MODELS

Annual data cover up movements occurring within the year. A model fitted to annual data contains less information than a quarterly model. The former will not yield quarterly information without many arbitrary assumptions, but a quarterly model can always be converted to an annual basis by summation.

To approximate an annual investment equation, we combine (3.2) and (3.6), putting ρ_2 = zero and using the values for other parameters given in (6.1). We obtain

$$\begin{aligned}(I)_3 + (I)_2 + (I)_1 + I &= A + 0.438(\Pi)_{3/2} + 0.414(\Pi)_{1/2} + 1.007(\Pi)_{-1/2}\\ &+ 0.953(\Pi)_{-3/2} + 0.346(\Pi)_{-5/2} + 0.327(\Pi)_{-7/2} - 0.184(\Pi)_{-9/2}\\ &- 0.174(\Pi)_{-11/2} - 0.198(K)_{-1} + U\end{aligned}$$

where A is a constant and U is a linear combination of disturbances which for the present purpose we may consider random.[25] Time zero is February 15 of the year for whose four quarters investment is estimated. Taking $(\Pi)_{3/2}$ and $(\Pi)_{1/2}$, together with one-half $(\Pi)_{-1/2}$ as relating to the current year, dividing the remaining lagged values of Π between the two preceding years, summing the coefficients, and dividing by 4, we obtain the following *annual* equation (time zero becoming June 30):

[24] At the time our model was estimated, no adequate breakdown of income taxes by kind of income was available. Annual estimates for disposable income have only recently been compiled in a fashion to segregate wage from other income: see Lenore Frane and L. R. Klein, "The Estimation of Disposable Income by Distributive Shares," *Review of Economics and Statistics*, Vol. XXXV (1953), pp. 333–37.

[25] To derive this result we substitute $[K - (K)_{-1}]$ for I in (6.1) and compute $[K_3 - (K)_{-1}]$ by recursion starting out with the expression for K, obtaining $(K)_1$, then $(K)_2$, and then $(K)_3$.

$$(7.1)\quad I = A + 0.34\Pi + 0.51(\Pi)_{-1} - 0.07(\Pi)_{-2} - 0.20(K)_{-1} + U.$$

This compares with the following equation obtained directly from annual data:[26]

$$(7.2)\quad I = A + 0.23\Pi + 0.55(\Pi)_{-1} - 0.15(K)_{-1} + U.$$

The long run interpretation of the quarterly consumption equation (6.3) has already been discussed. To convert this conveniently to an annual basis, write $(C)_{-3/2} = \frac{1}{2}[(C)_{-1} + (C)_{-2}]$. After some substitution and rearrangement the following is obtained (lags refer to quarters):

$$\begin{aligned}(C)_3 + (C)_2 + (C)_1 + C \\ &= A + 0.607(C)_{-1} + 0.877[(C)_{-2} + (C)_{-3} + (C)_{-4}] \\ &\quad + 0.270(C)_{-5} + 0.035(Y)_3 + 0.051(Y)_2 + 0.076(Y)_1 \\ &\quad + 0.096Y + 0.061(Y)_{-1} + 0.044(Y)_{-2} + 0.020(Y)_{-3} \\ &\quad + U.\end{aligned}$$

where as before A is a constant and U is a linear combination of variables that may be considered random. Summing coefficients applicable to respective years and dividing by 4 we obtain as an *annual* consumption equation:

$$(7.3)\quad C = A + 0.81(C)_{-1} + 0.07(C)_{-2} + 0.06Y + 0.03(Y)_{-1} + U.$$

The autoregressive component, although not quite so overwhelming as in the quarterly equation from which it is derived, still is far stronger than that obtainable directly from annual data:[27]

$$(7.4)\quad C = A + 0.46(C)_{-1} + 0.51Y + 0.03t + U.$$

The consumption equation in Model B may also be adapted to an annual basis. A procedure strictly analogous to that used with equation (6.3) when applied instead to (6.7) yields the following *annual* consumption equation:

$$(7.5)\quad \begin{aligned}C = A &+ 0.12(C)_{-1} + 0.01(C)_{-2} + 0.89(W_1 + W_2) \\ &+ 0.20(W_1 + W_2)_{-1} - 0.32\Pi - 0.07(\Pi)_{-1} + U.\end{aligned}$$

The closest comparable equation obtained directly from annual data reports positive coefficients for Π and $(\Pi)_{-1}$:[28]

$$(7.6)\quad C = A + 0.80(W_1 + W_2) + 0.02\Pi + 0.23(\Pi)_{-1} + U.$$

[26] Klein, *op. cit.*, p. 68.
[27] Carl Christ, "A Test of an Econometric Model for the U.S.," *Conference on Business Cycles*, p. 75.
[28] Klein, *op. cit.*, p. 68.

The two wage equations yield the following *annual* versions. From (6.5):

$$(7.7)\quad W_1 = A + 0.65(C + I + G - W_2) - 0.07(C + I + G - W_2)_{-1} + 16t + U.$$

From (6.8):

$$(7.8)\quad W_1 = A + 0.63(C + I + G - W_2) - 0.04(C + I + G - W_2)_{-1} + 45t + U.$$

Computed from annual data:[29]

$$(7.9)\quad W_1 = A + 0.42(C + I + G - W_2) + 0.16(C + I + G - W_2)_{-1} + 130t + U.$$

It will be seen that the sum of the coefficients of $(C+I+G-W_2)$, this year and last year, is practically the same in all three equations. However individual coefficients differ somewhat, and the time trend reported from annual data is much stronger than in either of our quarterly models.

CONCLUSION

To the question whether it is possible satisfactorily to represent quarterly movements of gross national product in the U. S. economy by as simple an equation system as that discussed here, these results offer no conclusive answer.

An inspection of the two models described reveals three main weaknesses. (1) In each model at least one of the three equations showed significant autocorrelation of residuals. (2) In the recursive model (A) the coefficient of income in the consumption equation is small, and its significance could not be established, so that the linkage between the first two equations in this model is poor. The hybrid model (B) shows a different, though probably related weakness—the coefficient of non-wage income in the consumption equation is negative. (3) In both models the sampling errors of some coefficients are uncomfortably large.

Constructed from data for 1923–1940, the models were tested by their performance during 1947–1952. (1) In a series of short-range predictions one quarter ahead both models performed uniformly better than guesswork, but their superiority was not decisive in a statistical sense. Probabilities against results being obtained by chance ranged

[29] *Ibid.*

from about 10 to 1 to much higher and clearly significant odds, depending upon the extent to which successive predictions were assumed to be independent. (2) In a single long-range prediction by each model for the entire six-year period, the econometric models forecast GNP with a smaller root mean square error than any of the guesswork schemes. In terms of components of GNP it is not possible to say that the models did appreciably better than guesswork.

Evidently room for improvement is large. The obvious advantage of more plentiful observations, both in constructing and testing of models, when quarterly data are used will not be fully realized until further progress has been made with the autocorrelation problem. It may well be that some of the defects of the models discussed can be overcome only through the use of a more detailed and complex system of equations.

APPENDIX

The equations for obtaining least-squares estimates of the parameters α_i of an equation with autocorrelated disturbances

$$x = \alpha_0 + \alpha_1 z_1 + \cdots + \alpha_n z_n + u \tag{8.1}$$
$$u = \rho(u)_{-1} + v \qquad (v \text{ random})$$

are obtained by choosing that set of estimates of α_i and ρ which minimize $\sum_1^T v^2$. The sample observations cover the period $t=1, 2, \cdots T$. We can rewrite (8,1):

$$x - \rho(x)_{-1} = \alpha_0(1 - \rho) + \alpha_1[z_1 - \rho(z_1)_{-1}] + \cdots + \alpha_n[z_n - \rho(z_n)_{-1}] + v. \tag{8.2}$$

(It will be noticed that if $\rho=1$, the case is equivalent to the use of first differences.) Assume that all variables are measured from their means, and write

$$X = x - \rho(x)_{-1}$$
$$Z_i = z_i - \rho(z_i)_{-1}.$$

Minimization of $\sum_1^T v^2$ with respect to α_i then leads to the familiar equations

$$\sum_1^T (XZ_i) = \alpha_1 \sum_1^T (Z_1 Z_i) + \cdots + \alpha_n \sum_1^T (Z_n Z_i) \tag{8.3}$$
$$i = 1, 2, \cdots n.$$

We proceed to minimize $\sum_1^T v^2$ with respect to ρ and obtain

$$(8.4)\quad \sum_{1}^{T} [X - \alpha_1 Z_1 - \cdots - \alpha_n Z_n]$$
$$[-(x)_{-1} + \alpha_1(z_1)_{-1} + \cdots + \alpha_n(z_n)_{-1}] = 0.$$

The sums of (8.3) are all quadratic expressions in ρ, so that we can solve for α_i in terms of ρ. The solution is in each case the ratio of two polynomials, each of degree $2n$ in ρ. The coefficients are throughout combinations of moments of sample observations. Inspection of (8.4) shows that the highest order terms will be of the form $\alpha_i\alpha_j\rho$. Therefore substitution of the polynomials obtained from (8.3) in (8.4) will lead to a polynomial of degree $(4n+1)$ in ρ.

Although of high degree (17th degree in the case of our investment equation, which has four independent variables) these equations in ρ are not hard to solve numerically by iteration, because the relevant (and often the only real) root must lie in the interval $1 \geqq \rho \geqq -1$ at least for series where the end effect may be neglected. For if u satisfies

$$u = \rho(u)_{-1} + v$$

then

$$Eu(u)_{-1} = \rho E(u)^2_{-1} + Ev(u)_{-1}.$$

Now $(u)_{-1}$ can be expressed as a constant plus a linear combination of past values of v up to $(v)_{-1}$. Hence if the v are mutually independent, $Ev(u)_{-1}$ must vanish and

$$(8.5)\qquad \rho = \frac{Eu(u)_{-1}}{E(u)^2_{-1}}$$

which is the autocorrelation coefficient for u in a long series.

Once (8.4) is solved for ρ, $\hat{\alpha}_i$ may be obtained by substituting $\hat{\rho}$ in (8.3).

If the residuals in the structural equation are assumed to satisfy a second or higher order difference equation, estimation is a much more difficult matter, since simultaneous nonlinear equations must be solved.

PROBLEMS OF COORDINATING THE UNITED STATES STATISTICAL SYSTEM

STUART A. RICE
U. S. Bureau of the Budget

THE statistical system of the United States embraces a great many official, semi-official, and unofficial agencies and instruments. Together these "comprise a *system* in the same sense that the activities of four and one-half million business units comprise a national economic system."[1] Elsewhere I have defined the terms here discussed as follows: "A statistical *system* exists when coherence is established and maintained among the separate programs that compose it. Such coherence requires an item-to-item adjustment of each task and process to every other related task and process, whether the relationship be one of conceptual congruity or one of consistency in operational patterns and sequences. The process of attaining and maintaining this coherence is called 'co-ordination'."[2]

The end sought is a better integration of our nation's statistical intelligence. Why is this desirable? I suggest two closely related reasons: First, an integrated system is more efficient. Second, it gives us a better understanding of the world in which we live.

Our world is precariously balanced between forces which are furthering advances in civilization and others which are pushing us toward universal catastrophe. The balance among them can easily be tipped in one direction or the other. We cannot afford to blunder because of an inadequate understanding of the forces with which we deal. Nor can we afford to spend a single taxpayer's dollar to lesser advantage than we might, when the demands upon our Government and upon the whole economy are so numerous and so pressing. If spent for statistics, that dollar should produce the greatest possible yield of useful information.

The needs of users of statistics are seldom limited to a single series. For example, they may need to know simultaneously the facts about employment and production—not separately but in relationship. The data used, whatever their separate sources, should intermesh. Employment and production series will not intermesh unless the definitions of reporting units correspond and unless they are grouped in accordance

[1] Stuart A. Rice, "The Role and Management of the Federal Statistical System," *The American Political Science Review*, XXXIV (1940), p. 481.

[2] *Ibid.*, "Co-ordination of Federal Statistical Programs," *The American Journal of Sociology*, I (1944), p. 22.

with the same industrial classification. Not so long ago, as the life of a bureaucrat goes, both "employment" and "industries" were separately and often inconsistently defined and classified by different Federal agencies. Hence, basic steps toward the integration of our system of statistical intelligence were taken with the development, promulgation, and incorporation into general use of the standard industrial classification. Other steps were the standardization of the types by status of persons included in the economically active population and the establishment of uniformity in reporting periods for employment.

The importance of such standards is illustrated in reverse when, in ignorance of them, investigators of economic and social problems formulate their own categories. Too late they may discover that their results are noncomparable with basic data, like those of the Bureau of the Census, with which to have significance they should be aligned.

The relative merits of two general methods of integrating a statistical system have long been debated. One is administrative centralization, achieved with such conspicuous effectiveness in Canada. The other is decentralization, accompanied by coordination, exemplified by the statistical system of the United States. After many discussions of these assumed alternatives it is my judgment that the issues between them are largely unreal. In any event I disclaim partisanship. The statistical system of any country is an outgrowth of historical development. Within it are strains toward adaptation to the political, social, and economic structure which it serves and of which it is a part.

The United States has developed an over-all pattern of statistical decentralization from which it is now too late to depart. Federal statistical activities began to proliferate in a decentralized fashion immediately after the birth of the republic and the trend thus established has continued. The pattern is remarkably adaptive to its *milieu*. The uniquely large volume of information produced by the statistical system of the United States reflects the factual-mindedness of our people. Statistics have been in demand and the demand could not have been satisfied so easily or so well except under the conditions provided by the historical decentralization of the nation's statistical mechanisms.

Other considerations support the efficacy for us of our own decentralized system. It permits a useful division of labor among agencies, public and private. It keeps the collection of many data in close contact with the uses to which they are to be put when assembled, thus giving protection and assurance to administrators in the fields of utilization.

Nevertheless, without a central mechanism for the coordination of statistical activities the scene presented in a decentralized statistical

system would be chaotic. In this country the central agency of statistical coordination, known as the Office of Statistical Standards, is a part of the Bureau of the Budget in the Executive Office of the President. Its legal powers, affecting other Federal agencies, are more authoritative than it usually cares to assert; and its interests and concern, though not its legal authority, extend to numerous statistical activities which are under private direction.

The limitation of the formal authorities of the Office of Statistical Standards to Federal agencies does not impede the integration of the nation's statistical system. True, much statistical work is carried on by trade organizations and business units which are highly competitive. However, the dominance of the Federal Government in statistics and the prestige and influence of such Federal agencies as the Bureau of the Census and the Bureau of Labor Statistics are so great that business-produced statistics are coordinated informally to a very substantial extent.

Within the statistical system the problems of coordination are numerous and I shall have to offer a selection.

1. The first is the problem of representing the general interest. It is naive to assume that governments are monolithic. The definite article before the noun—"*the* Government of the United States"—reflects an ideal that does not fully exist in actuality. The ensemble of Federal statistical agencies is not dissimilar to a trade association. The individual Federal agencies, like the individual members of a trade organization for an industry, are competitive but recognize many common interests. To carry the analogy farther, the Office of Statistical Standards might be regarded as the secretariat of the trade association which serves the Federal statistical products industry.

Customary procedures for financing and administering the separate members of this Federal statistics trade association tend to emphasize the autonomy of its members. The preparation of estimates of expenditures, their defense before the Bureau of the Budget and committees of the Congress, and the administration of approved programs are all responsibilities of the separate agency units to which funds are specifically appropriated. These procedures make statistical coordination more difficult, for *an integrated statistical system should serve public and governmental interests which override or lie between the interests and responsibilities of particular agencies.*

Even when a Federal agency is persuaded to include in its estimates items of expenditure for purposes beyond its direct responsibilities, such items are inevitably the first to suffer when reductions are made,

whether in the budgeting, the appropriating, or the administering stages. The OSS recognizes an obligation to serve as a "public defender" for general interests in statistical data; but recognition of this obligation in the Congress and even within the Administration itself is very limited. Neither specific legal authorities nor the sanctions conveyed by custom are present to lend weight to its appeals. When it has "gone to bat" in public for comprehensive objectives, as with the "reconversion statistics program" of 1945, the consequences have not been encouraging.

The reasons are not far to seek: First, it is contrary to tradition for representatives of the Budget Bureau to appear before committees of Congress *on behalf of* appropriations. Secondly, the Bureau's functions are properly inconspicuous and anonymous. They do not build up the type of public support which, in the case of other agencies, is sometimes mobilized behind programs.

Such considerations as these led the Mills-Long task force on statistics to recommend to Mr. Herbert Hoover's first Commission on Organization of the Executive Branch of the Federal Government that the OSS be put in possession of free funds for disbursement at its discretion on behalf of Federal statistical interests. This is not the place to outline the objections to this solution.

A start has been made in another direction toward accomplishing the same objective. Each year the Office of Statistical Standards prepares a so-called "statistical budget," recommending programs to be undertaken by the principal general-purpose statistical agencies: Bureau of the Census, Bureau of Labor Statistics, Bureau of Agricultural Economics, Office of Business Economics, National Office of Vital Statistics, and such joint enterprises as that on financial statistics of the Federal Trade Commission and the Securities and Exchange Commission. The purpose is to secure over-all balance and thus achieve a Federal rather than a series of unrelated departmental programs. This "statistical budget" is reviewed in the Bureau of the Budget like any other Federal proposal.

The "general interest" has also been represented and furthered by the Office of Statistical Standards in various other ways. I have already mentioned some of the standard classifications and definitions that it has developed cooperatively with the "operating" statistical agencies for the use of all of them. For the Joint Committee on the Economic Report of the Senate and House of Representatives and for the Council of Economic Advisers it has prepared analyses of the gaps in our national arsenal of statistical information. It was our privilege some

years ago to initiate the development of the monthly publication *Economic Indicators*, now prepared for the Joint Committee by the Council; and we have just completed the technical work upon a supplement which interprets the sources from which the indicators are derived. Our "Blue Book" on *Statistical Services of the United States Government* has had wide national and international circulation. Our *Federal Statistical Directory* and our monthly *Statistical Reporter* have been useful instruments of statistical integration within the Government. Not least important in this list are the facilities we offer for consultation upon the general interest by the various Federal statistical agencies, coming together at our invitation upon neutral ground.

2. Certain consequential problems are involved in the procedure of developing a statistical budget. The separate items it brings together must also find a place in the estimates of expenditure submitted and supported by the respective departments and agencies which will administer the funds. Hence the statistical items must be adjusted to other items within the departmental budgets concerned. Occasions have arisen in which the Bureau of the Budget has differed with the head of a department or agency concerning priorities among the statistical and other items in his budget. There have been cases in which funds approved by the Bureau for statistical purposes have been diverted by administrators to other purposes which they deemed more important. Their actions could not easily be challenged without violating the sound principle that the responsibilities of an administrator should be accompanied by command of the resources given him.

3. Almost inseparable from the problem of defending the general interest is that of securing balance among particular agency and other interests of which a coordinated statistical program must take account. Each Federal agency wishes to collect data for which it feels an administrative need or for which there is a legal or public demand imposed upon it. However, through processes of coordination, a single inquiry may often be made to serve additional purposes. The data produced can sometimes supply the essential needs of other agencies as well. The original agency is not necessarily made happy by this prospect. It is exposed to the danger that its own purposes may be inadequately or belatedly advanced.

Dangers to the "other agencies" are also present. The relationships established when one agency renders service to another may introduce seeming exceptions to the principle that responsibility should be linked with command. The relationships acquire a contractual character, especially if a regular flow of data from the servicing agency to its

"customer" agency is desired. In scheduling production the administrator of the first must resist the frequent temptation to subordinate the interests of the second. There is perhaps no greater impediment to statistical integration than the fear that if important statistical work is contracted out to another organization it will lose priority.

The attainment of optimum balance in an omnibus program, giving appropriate weight to each separate interest and to the general welfare, is one of the most delicate problems of statistical coordination.

4. Another difficult problem is that of establishing demarcation lines between governmental and nongovernmental responsibilities for the collection of statistics. The governing principle is clearly that data should be collected at the expense of government only when vested with public interest. This formula shares the simplicity of the well known secret of money making in the stock market—buying low and selling high. The difficulties appear in the application of the principle. Few statistical series on any subject are without *some* public importance. Even a strictly *private* interest, if shared by a sufficient number of people, becomes of public interest through its implications for the economy. Aids to farmers in marketing their crops provide examples. To what figure would the number of beneficiaries have to be reduced for the "public" interest in statistical estimates of crop production to become strictly "private"? Ten thousand? One thousand? One hundred? Ten? Or the single producer?

In historical fact, conceptions of what may or may not be appropriate undertakings by government are under constant revision. I have tried for many years to find some magic formula by which to apply the criterion of public interest to governmental statistical activities. I conclude that there is no general criterion and that the conception must be applied to individual situations as they arise, instance by instance.

5. The extent to which the protection of "confidentiality" should be thrown around industrial or company data supplied to an agency of government raises almost equally difficult problems of discrimination. These evoke the emotions of respondents; they provoke disputation between exponents of monistic and pluralistic conceptions of government; and they produce headaches for a statistical coördinating agency.

Perhaps I am guilty of some bias in favor of the pluralistic conception when I say that once again the governing principle is clear. This is that data supplied to an agency of government for statistical purposes should not be allowed, through disclosure, to cause individual

hardship or disadvantage. It should not be used to support legal action against the respondent in the courts. It should not fall into the hands of business competitors who would find therein a competitive advantage.

By becoming an unreasoned *fetish* this wholesome principle has actually worked in the past to considerable disadvantage to both government and suppliers of statistical information. For many years two Federal statistical agencies, each legally empowered to collect the same information from the same respondents, believed themselves unable to share this information with each other and thus avoid the necessity of duplicating their inquiries of the public.

This situation was mitigated by certain of the provisions of the Federal Reports Act of 1942; yet the "fetish" remains as an impediment to many practical steps that would otherwise assist in the integration of Federal statistical activities. In a recent instance agency A sought to avoid the dilemma by asking its business respondents to supply information previously given to agency B, in accordance with specifications copied by A from those of B. Still more recently it was demonstrated that some millions of dollars might be saved in conducting the (ill-fated) Census of Business for 1953 by utilizing the information reported on income tax returns by retailers having no employes. The procedures proposed would have provided full protection against disclosure to competitors or to Federal agencies other than the two directly concerned. When the proposal was considered by an advisory group of business consultants it was initially viewed with a skepticism approaching horror. Gradually the motive of economy prevailed and its adoption was recommended.

Encroachments by governments upon the liberties of individuals pose one of the great politico-ethical problems of our day. Federal statisticians and respondents alike are handicapped by the absence of a clear understanding and agreement upon the limits of "confidentiality" that should be attached to statistical returns.

6. Analogous and sometimes related questions are raised by the needs of government to withhold from its own citizens certain statistical data which, if they reached potential enemies, would give aid and comfort to the latter. Since these questions are not inherent in the processes of statistical coordination *per se*, I shall leave them aside with a single footnote reference.[3]

7. Lastly I mention the need for liaison between Federal statistical

[3] Stuart A. Rice and Joseph W. Kappel, "Strategic Intelligence and the Publication of Statistics," *The American Political Science Review*, XLV (1951).

agencies and the statistical profession. How can we in the Federal Government best consult with our nongovernmental colleagues? How can we obtain their advice upon the issues we face—advice which is at the same time technically competent, and fully informed respecting the setting in which the issues arise? The last condition is essential if advice given us is to be realistic. If the condition is met, the preparations for advice-giving must of necessity be very time-consuming, both for advisors and advisees.

In the lucid and "sobering" analysis of the responsibilities that have been placed upon Federal statistics in connection with the nation's practical affairs, presented in her notable Presidential address at Chicago in December, 1952, Mrs. Wickens[4] grappled courageously with this thorny problem. She felt that the time had come "for the profession as a whole to share some responsibility for these statistics with those who make them." She therefore proposed "that there be created a new United States Statistical Commission, with responsibility for audit of statistical series, similar to an accounting audit, empowered to put a 'certified' label on a statistical product. It should also be charged with investigation of methods, scope, and suitability of statistics, and with making recommendations for future improvements and developmental work. . . . Primarily, its membership would be drawn from experts outside government. . . . It should be a continuing body, serving on occasion as required, but with a small full-time staff, and adequate financing, so that our most distinguished statisticians, economists, scientists, and other specialists could reasonably be expected to devote time and attention to its work. . . ."

Mrs. Wickens' proposal excited admiration for its boldness and breadth. It remains my opinion that the objectives which she visualized can be realized in the present only piecemeal and on a much more modest scale. How would the Commission, as a "continuing body," be financed? If through Federal appropriations, how long could it escape the constricting influences that surround other Federal agencies? How would it divide its time between its technical functions and the annual and prolonged struggles over appropriations, over personnel appointments and security clearances? Presumably such an organization would have to operate under the full and specific authorities possessed by the Office of Statistical Standards; but, how would it become related to such other existing Federal agencies with central functions and legal authorities as the Council of Economic Advisers?

[4] Aryness Joy Wickens, "Statistics and the Public Interest," *Journal of the American Statistical Association*, 48 (1953), pp. 1–14.

Most difficult of all, where would it find the distinguished members who could devote the requisite time to the labors that are visualized? Whatever the competence of the individuals who compose them, high level committees and commissions need continuity of membership and long exposure to the problems upon which they are asked to advise. The inevitably complex problems of government statistical organization cannot be quickly grasped. The Commission would demand members having a high degree of competence and general experience, but such people are invariably very busy.

I can vigorously applaud Mrs. Wickens' references to the "first steps" made by our profession "in these directions." I feel great satisfaction in the services given to the Office of Statistical Standards by the Advisory Committee on Statistical Policy, created by the American Statistical Association at our request in 1951 and composed of members appointed from a distinguished panel of the Association's present and past presidents and president-elect. Methodically, and without undue haste, the Committee has thought its way through the intricacies of a number of perplexing policy issues regarding which the Federal statistical services are entitled to look for leadership to the Office of Statistical Standards. These have involved some of the problems that I have discussed above, including that pertaining to the confidentiality of individual returns and differentiation between the appropriate areas of governmental and nongovernmental agencies in the collection of data.

The Committee is slowly approaching some of the other tasks which Mrs. Wickens proposed for the new Statistical Commission, such as "making recommendations for future improvements and developmental work." I cannot foresee the arrival of a time in its work when it could itself undertake "responsibility for audit of statistical series" or detailed investigations of "methods, scope and suitability" of Federal statistics.

We should like to make it possible, through suitable compensation for the services it renders, for the Advisory Committee on Statistical Policy to render even greater service in the future than it has in the past. We do believe that it should avoid becoming entangled in small issues and that it should be able to sift out for its attention policy questions of the highest priority; for we do not want a "captive committee" that can be presented at regular intervals with a long list of activities for its approval.

Other types of advice are needed and received by the Office of Statistical Standards from other sources. Intra-governmental com-

mittees and conferences of representatives of Federal statistical agencies are meeting daily upon specific questions of coordination. The Advisory Council on Federal Reports, representing the world of industry and business, operates with its own budget and secretariat, bringing to bear upon us the viewpoints and interests of respondents upon the procedural problems which arise in the Federal collection of data. Similarly, the Labor Advisory Committee on Statistics presents to us the needs for Government figures of an important segment of the statistics-consuming public.

Mrs. Wickens closed her memorable address with the expression of belief that "As a profession, statisticians must organize to meet this challenge if statistics are to continue to be administered in the public interest." Despite many set-backs and discouragements, and in ways less pretentious than by the specific device she proposed, I believe they are doing so.

GROWTH BY MERGER*

G. Warren Nutter
Yale University

Like the business cycle in the recent past, industrial organization is a topic more talked about than investigated. Every new empirical study will be devoured by the fact-hungry specialist in this field, in the hope that it will answer some of the long list of unsettled questions. Professor Weston's recent book on mergers will surely be avidly read, but it will not relieve the hunger as much as one would hope.

The book deals with many things, although about half of it is concerned with the main topic, namely, the role mergers have played in the absolute and relative growth of large firms. The remaining half discusses the particulars of the most recent merger movement, the theory of mergers, and problems of economic policy. This review will concentrate on the main topic: it is of greater interest to the general economist than the others and it lies less completely outside my areas of competence. The remainder of the book is certainly worthy of attention; it is slighted here because I find it less controversial and should have little to contribute on it in any event, in view of my bare nodding acquaintance with the history of mergers. The chapter on mergers of the 1940's is a useful and interesting summary of the more reliable studies of this movement. The chapter on the theory of mergers, though it contains some doubtful conclusions on the motives behind them, sheds new light on factors explaining the timing of merger movements. The chapter on economic policy adds few original suggestions but summarizes the problems rather well. I leave it to experts on the history of mergers to examine the details of these chapters.

THE CENTRAL ISSUE

As I read the history of economic controversy, the central issue about mergers is whether industrial concentration and corporate giantism are in any significant degree traceable to them. There are two other subsidiary issues: whether concentration and bigness are the usual results of mergers; and whether they are the primary goals. These are all important issues, but they must be recognized as distinct. Though the point will not be developed further, it can be noted that Weston does not always avoid mixing them up.[1]

* A review article on J. Fred Weston, *The Role of Mergers in the Growth of Large Firms* (Berkeley and Los Angeles: University of California Press, 1953). Pp. xvi, 159. $3.50.

[1] See particularly the discussion in his book on pp. 34–37 and 51–57.

The first issue is of importance because an answer to it may cast light on the bases of monopoly (or, to use a more widely accepted word, oligopoly) and big business. Theory tells us that a firm may achieve a dominant position in an industry for any of three broad reasons: (a) it may have genuine economies of scale relative to the market it operates in; (b) it may have "artificial" economies, such as patents; or (c) it may be able to fare better than average over the long haul through temporary exploitations of its dominant position. It is virtually impossible to discover the comparative importance of these factors by direct searching for causes; we cannot, for instance, measure economies of scale. Hence we must rely on inferences from other, more indirect, evidence. One relevant matter is the way in which firms have been put together. It is at this point that the question of mergers enters. If a dominant firm has chosen to achieve a significant part of its growth through mergers, doubt is cast on the importance of economies of scale, certainly as they might derive from plant operations. The greater the extent to which dominance has been maintained through mergers, the less likely that economies of scale are a general basis of monopoly (or oligopoly), particularly if the typical picture is one of a continual struggle to retain dominance in the face of continual encroachment by other firms.

The link between analysis of mergers and bigness is similar. Without, of course, exhausting all possible reasons for corporate giants, one may suppose that the most important are advantages of monopolistic position and economies of producing multiple products. Here again the method and nature of growth helps us to discover reasons, if only by process of elimination.

It is hard to know if Weston shares this view of underlying issues. In the opening sentences of his book he does suggest that interest in mergers arises from concern over broader questions, but he does not get down to details. He formulates the empirical problem in very general terms, apparently in order to make the findings applicable to a wide range of specific issues. "The sources, extent, and consequences of bigness and concentration have been widely disputed," he says, and "many questions remain unsettled." He goes on to list some of them and concludes that, "although the formation of appropriate public policies awaits further study of these and related issues, our understanding of the nature and appropriate role of large firms may be increased by more complete information concerning the process of their growth" [7, p. 1].

There is little to dispute here, if we substitute "relevant" for "complete," find out what is relevant, and frame the empirical problem accordingly. Instead of using this approach, however, Weston looks di-

rectly to economic literature and concludes that "mergers are often cited as the major source of economic concentration" [7, p. 2]. Hence he sets this up as the empirical question to be resolved.

Some economists have undoubtedly made statements like this,[2] but it is doubtful that such statements reflect the heart of the controversy over the role of mergers. In any event, Weston's framing of the empirical question is unfortunate; for, as his later analysis reveals, he is trying to find out whether mergers account for more or less than half of the growth of "oligopolistic" firms, a matter of limited relevance to the underlying issues as I see them. The real question is whether the pattern of concentration and bigness as we now observe it would have emerged in the absence of mergers; that is, whether mergers have accounted for a *significant* portion of growth. There would seem to be no reason for choosing the fraction one-half as the dividing line between significance and insignificance. Yet he does so, explicitly as well as implicitly. For example, he characterizes mergers as a negligible factor in growth when he finds that they have accounted on the average for a fourth to a third of the absolute growth of firms studied [7, p. 30].

It would do Weston's work an injustice to say that his data bear in no way on central issues; on the contrary, there is much here with direct bearing, available for the first time. I lament that there is not more, for there could have been with a slight change of orientation. Weston has collected certain types of data. He has used only a portion of these, summarized in ways best suited to his objectives. He has not given much attention to the problem of ordering the basic data in a flexible manner, a defect made more acute by failure to publish many of them. All of which is to say that it is very difficult to make much additional analysis of the data presented. On the other hand, something can be said about Weston's findings, which is the task we turn to next.

ABSOLUTE GROWTH OF FIRMS

Weston selects 22 census industries with highly concentrated output in 1935 and studies the growth of 74 of their dominant firms,[3] beginning in the earliest year for which data could be found in each case and end-

[2] But not in all of the places Weston cites. For instance, Corwin Edwards does not say anything about the quantitative importance of mergers [1, p. 142]; Stigler says something different [4, p. 23], as pointed out below in this review; and the statement from a government report quoted by Weston says that mergers are "probably more important than any other single factor" [7, p. 2], again not the same as Weston's statement.

[3] He explains the basis for selecting industries [7, pp. 112–21] but not firms, other than to say that the latter are dominant in the selected industries. Important firms are omitted from several industries, e.g., electrical machinery, ammunition, ink, sewing machines, and cement. One is led to suppose that ease in obtaining data played a part in selection of the sample.

ing in 1948. Growth is measured by accretion in value of total assets.[4] Growth by merger, or "external" growth, is defined as the accretion resulting directly from all types of acquisitions of already existing firms; the residual is viewed as "internal" growth, that is, as growth by means other than merger. The relative importance of external growth is measured by the fraction of total growth accounted for by mergers, which we shall call "proportional growth by merger." Because of conceptual and empirical difficulties in tracing external growth to its beginnings, Weston develops three measures, differing from each other in their treatments of assets in the earliest year for which he could find data. In the first measure the initial assets are regarded as external growth, that is, as resulting entirely from mergers; in the second they are excluded from both external and total growth, only subsequent growth being measured; in the third they are excluded from both external and internal growth but not from total growth—that is, they are considered a separate component of growth. The first measure gives the highest estimate of proportional growth by merger, the third gives the lowest, and the second gives an intermediate estimate.

Weston shows no preference for one measure over the other, and one must admit that it is difficult to choose among them. The first would be clearly the best if the initial assets of the largest firm taking part in the earliest merger were not counted as external growth, though there are conceptual problems here, too.[5] As the measures stand, one can say that the second and third, taken by themselves, are more misleading than the first, because their understatement magnifies a strong downward bias that is present for another reason, discussed below. It is my guess that this bias overwhelms all others; therefore, at least for judging the over-all importance of mergers, the first measure would seem to be least bad, on the grounds that it probably minimizes a general downward bias.

Like all empirical workers, Weston was faced with a host of measurement problems, none having a wholly satisfactory solution. He discusses most of these quite adequately. For instance, he devotes an appendix to the question whether total assets are a better index of size than any component of assets. He divides the basic data on firms into three groups, on the basis of the reliability that he attributes to them.

---

[4] Growth of firms in the steel industry is also sometimes measured in terms of tons of ingot steel capacity [7, pp. 22–23 and 132–34]. The reason for this measure, and for selected use of it (not always called to one's attention), is not explained.

[5] Weston does something similar to this in an isolated appendix table [7, pp. 151–52], which plays no part in his analysis. He does not explain why the procedure was followed here or why it was not adopted as a general practice.

According to his judgments, data are "accurate" (were fully confirmed by companies) for firms accounting for about a third of aggregate assets in 1948; "dependable" (partly confirmed) for firms accounting for about a half; and "questionable" (unconfirmed) for firms accounting for about a sixth. In general, he segregates findings for each group so that some allowance may be made for the relative shortcomings of basic information. Finally, he points out that his measures of external growth cover only that part of growth directly attributable to mergers; they do not take account of indirect effects. He argues that any effort to summarize both direct and indirect effects would lead to unanswerable questions about how growth would have gone in the absence of mergers. In some cases it might have gone worse, in others better. This view seems reasonable enough in terms of strict logic, and examples can be found of both effects. Nevertheless, one wonders whether there is not in general a presumptive advantage in mergers, at least as far as relative growth is concerned. Otherwise, why do firms continue to engage in them?[6]

In view of the attention paid to some of these problems, it is strange to find no mention of one that probably overweighs all others: how to deflate the value of assets to allow for changing price levels. Even if we waive all other troublesome problems connected with measuring growth of capital (which is in itself hardly permissible), we are not justified in taking a dollar's worth of assets as representing the same real value in, say, both 1900 and 1948; some adjustment must be made for differences in price levels. By Weston's procedures, total growth is represented by the value of assets recorded in 1948.[7] Thus it is measured in terms of recent prices. External growth, on the other hand, is represented by assets valued at the times mergers occurred. Since there has been a secular rise in relevant price levels, proportional growth by merger has been understated in this respect. Moreover, the understatement is probably of considerable magnitude; for, in spite of the understatement, Weston's data on all firms as a group show that a large share of growth by merger occurred in early years: 8 per cent before 1911, 21 per cent before 1921, 71 per cent before 1931, and 89 per cent before 1941 [7, pp. 155–56]. Unfortunately, the deflation problem is formidable.[8] There is simply no easy way to determine the degree of under-

[6] Weston suggests other reasons for mergers, applying primarily to recent ones [7, pp. 70–75]. But they seem to boil down to a list of ways in which mergers have an advantage over internal expansion.

[7] This is reduced in the second of his measures by the value of initial assets.

[8] This problem might have been partly avoided, or at least some notion of the likely bias might have been gotten, by measuring growth in terms of employment. The suggestion is made, however, with no knowledge of the empirical difficulties involved.

statement, particularly on the basis of the data presented. The best we can do is to keep this qualification in mind when examining Weston's findings.

For the firms as a group, external growth accounted for 33, 22, and 19 per cent of total growth, under the three measures used. The highest fraction occurs of course when initial assets are counted as external growth. For each measure the fraction is significantly higher for firms with "questionable" data than for other firms.[9] It is difficult to know how to interpret this, however, since the direction of error in the former is unknown. There appears to be no general relation between proportional growth by merger and size of firm; unweighted and weighted means are close together. At the same time, there is a wide dispersion among firms. For instance, when initial assets are counted as growth by merger, proportional growth by merger ranges from less than 3 per cent (Reynolds Tobacco) to almost 85 per cent (B. F. Goodrich). The median is about 30 per cent.

The same dispersion is apparent among industries. When the 74 firms are squeezed into 22 industrial categories, proportional growth by merger, expressed as the weighted mean for firms in each industry, ranges from 8 per cent for aluminum to almost 70 per cent for cement, ammunition, and steel.[10] The median is about 36 per cent. The data by industry are summarized in the table below. It might be noted that Weston does not give the weighted means in his tabular presentation [7, p. 22], and in some cases these differ rather markedly from unweighted means (for instance, asphalted-felt floor coverings, photographic apparatus, and dairy products). He also uses a figure for steel derived from growth of ingot steel capacity, a figure that is much lower than the one derived from growth of total assets.

It must be repeated that all figures are reduced if Weston's other two measures are used, some quite drastically. For instance, if initial assets are excluded from consideration, ammunition tumbles from the top of the list of industries (70 per cent) down to the bottom (2 per cent).[11]

What are we to conclude from this evidence? Let us see first what Weston concludes. "It appears," he says immediately following the presentation of evidence, "that as a group, and irrespective of measure-

[9] Weston considers the differences as "not great." Whether great or not, they must be taken as significant, running as follows: 41 per cent as compared with 36 per cent (for firms with "dependable" data) and 26 per cent (for firms with "accurate" data); 32 as compared with 23 and 18; and 27 as compared with 19 and 16.

[10] The figures for cement and ammunition are probably not very meaningful. Only two cement companies are covered, one (Lone Star) accounting for 7 per cent of output in 1945, the other (Ideal) for 3 per cent [7, p. 40]. Only one ammunition company (Remington Arms) is covered.

[11] But see n. 10 above.

ment assumptions, the firms studied achieved the major extent of their growth through internal development. The proportion of growth through external acquisitions, however, is appreciable" [7, p. 15]. This at one point; but later he says something quite different. In his sum-

PROPORTIONAL GROWTH BY MERGER IN 22 INDUSTRIES[a]

Industry	Unweighted Mean[b] (%)	Weighted Mean[c] (%)	Range (%)
Ammunition[d]	70	70	[d]
Cement[e]	69	69	67–72
Steel	64[f]	67	47–77
Compressed and liquefied gases	52	52	51–52
Asphalted-felt floor coverings	51	42	34–68
Typewriters and parts	51	54	8–74
Photographic apparatus[e]	49	28	12–85
Dairy products	48	62	14–82
Corn syrup, sugar, oil	44	49	15–66
Tin cans[e]	41	39	34–47
Liquors	40	39	27–59
Rubber tires	38	33	5–85
Meat packing	30	33	11–60
Petroleum refining	28	27	11–44
Cigarettes	25	21	3–58
Ink[e]	22	23	18–26
Rayon and allied products	22	31	10–39
Electrical machinery[e]	21	20	17–25
Motor vehicles	21	18	6–35
Sewing machines[d]	17	17	[d]
Agricultural implements	14	19	8–25
Aluminum	10	8	7–14

[a] Initial assets counted as growth by merger.
[b] Arithmetic mean of percentages for firms in each industry. Source: [7, p. 22].
[c] For all firms in each industry as a group, aggregate growth by merger as a percentage of total growth. Source: [7, appendix E].
[d] Only one firm covered.
[e] Only two firms covered.
[f] Weston gives 53 per cent, a figure based on growth of ingot steel capacity expressed in tons.

ming up at the end of the chapter on absolute growth he states: "Acquisitions have been a negligible portion of the total growth of most of the firms in census industries now characterized by a high degree of concentration in output. . . . The direct effect of mergers on the absolute size of large firms appears to have been small" [7, p. 30]. And in his final summing up: "The extent to which individual firms have grown

by acquisitions varies greatly, but external growth is a relatively minor fraction of the total growth of most of the firms" [7, p. 101]. There is a rather wide gulf between "appreciable," on the one hand, and "negligible," "small," and "relatively minor," on the other. There is little doubt from the general tone of his argument and widely scattered references that Weston really considers mergers a negligible source of growth. Yet his temporary wavering in the other direction is significant; in fact, it is the key for understanding why Weston adopts his final conclusion.

The point is simply this: As was said earlier, the question Weston sets out to examine is whether mergers have been "the major source of economic concentration." If "the major source" is understood to mean "responsible for substantially more than half," fractions below a half become minor, or small or negligible. It is now clear that Weston does conceive of the issue in these terms. Otherwise, why his conclusions? It is difficult to see any other grounds for calling a fraction of 33 per cent (or 22 or 19 per cent) negligible.

The impressive thing, to me at least, is the height of the fractions not their depth—particularly when account is taken of their probable downward bias and of the vast growth of the economy over the last half century, which leads us to expect internal growth to swamp growth by merger. Growth by merger has certainly been important enough, in a sizable group of the industries studied, to cast serious doubt on some of the explanations advanced for industrial concentration and abnormally large size.

Whatever conclusions are drawn, they should be considered tentative. Although this study makes a useful contribution, much remains to be done. It would be useful, for instance, to know the role played by mergers in industries that were highly concentrated around the turn of the century but are no longer; and in industries with continually low concentration. It would also be useful to study the changing importance of mergers over time. Weston unfortunately provides little information on this matter; the data presented by him are limited to broad sweeps of time ending in 1948, not broken down into subintervals. Finally, learning about the role of mergers in absolute growth only starts us on the way to learning about the role in relative growth. This leads us into the next topic discussed by Weston.

RELATIVE GROWTH OF FIRMS

Weston opens the discussion of relative growth by characterizing the trend in industrial concentration as a movement from partial monopoly around the turn of the century to oligopoly in recent times. The period

of partial monopoly (dominance of an industry by a single firm) is in turn described as a temporary diversion from a historically typical pattern of oligopoly (multiple dominance), as far as concentrated industries are concerned. That diversion he attributes almost entirely to the merger movement around the turn of the century. That is, growth by merger is in his view the primary explanation for increased concentration in specific industries in the 1890's and early 1900's. Later developments, he asserts, are a different matter: "acquisitions subsequent to the early merger movement have had relatively small effects on concentration" [7, p. 48].

Before we proceed to the core of his argument, a digressive comment on the early merger movement should be made. Weston's description of the result as a general trend from multiple to single dominance of industries implies that market areas remained essentially fixed during the latter part of the nineteenth century. This was, however, a period in which the truly national market was emerging; localized markets were much more prevalent in the years leading up to the mergers than thereafter. It is therefore not unreasonable to view mergers as a force counteracting expansion of markets, and hence as leading to less change in the structure of dominance than is usually believed. This point is raised not to refute Weston's argument, but rather to shift the line of argument in a rather obvious way. The point is this: The early merger movement is important in the history of industrial concentration because it made concentration, taken in a relevant sense, greater than it otherwise would have been, irrespective of whether it actually increased concentration or not.

Now, when we come to developments after 1904, the primary issue must be put in a similar way: Have mergers played an important role in making the pattern of concentration significantly different from what it otherwise would have been? Weston raises this question, but only after he is far along in his discussion; and then it is put as more or less subordinate to two other, narrower, issues. First, he wants to know whether the trend from single to multiple dominance can be attributed to a change in the nature and motivation of mergers. This question is essentially part of a running argument with Professor Stigler, on which comment will be deferred until later. Secondly, he wants to know whether mergers since 1904 have been accompanied by increased concentration. He concludes that they have not. It is in qualification of this conclusion that he raises the central issue, namely, "decreases in concentration might have been even larger in the absence of mergers" [7, p. 44].

In order to understand Weston's analysis of the central issue, we are almost forced to follow his rather wandering path to it. As he sets out to trace the relation between mergers and trends in industrial concentration, he is immediately confronted with the frustrating lack of reliable historical data on industrial concentration. Every one who has struggled with this problem will feel sympathy for Weston. Concentration ratios can be compiled from the Census of Manufacturers from 1929 to date; many already have been. However, ratios for individual firms are not available even here, since the data are grouped for no fewer than the four largest firms in each industry. Moreover, ratios for different years are not strictly comparable because of different definitions of industries. A fairly large fund of information can be gathered for the turn of the century, much of it broken down by individual firms;[12] but its reliability is doubtful, to say the least. The investigator is left to his own devices in constructing time series; he must search trade journals, isolated monographs, and so on. The difficulties can scarcely be exaggerated.

Weston pulls together estimates of each leading firm's share of output in 9 of his original 22 industries, covering selected years over the last half century. The estimates for 5 industries (motor vehicles, steel, cigarettes, aluminum, and cement) are as reliable as one can expect, though they could be more complete. On the other hand, the estimates for the remaining 4 industries (electrical machinery, meat packing, rubber tires, and tin cans) are built on a very shaky foundation, and it doubtful that much significance can be attached to them. In these latter cases, each firm's share of output is taken to be the same as the ratio of its sales of all products to the value of products for the census industry in which that firm can be classified. The size of probable errors under this procedure is so large as to destroy the significance of all but very large differences in shares at different dates.[13] It must be granted that census value of products for an industry, when computed on an establishment basis, includes the value of some products not classified in the industry; but it would be highly unlikely that errors here and in sales data for firms would be compensating, among firms and between firm and industry. In a footnote to his table Weston recognizes that the data "are not strictly comparable," but he believes

[12] See, e.g., [2, pp. 129–40].

[13] Some indication of possible errors is provided by comparing the combined share of the four leading firms for around 1935 as derived from Weston's estimates, in some cases by interpolation, with the combined share for 1935 computed from census data. For meat packing, the two are 30 and 50 per cent, respectively; for rubber tires, 73 and 81 per cent; for tin cans, 87 (two firms), and 80 (four firms) per cent; and for electrical machinery, 37 (two firms) and 44 (four firms) per cent [7, pp. 40–41 and 116].

that "the percentages through time indicate very roughly trends in occupancy of the market" [7, p. 40]. Even this limited conclusion is open to serious doubt. Moreover, in his analysis he treats changes in shares of output, measured in this way, as having much more accuracy than would be possessed by rough indicators of trends.

Weston's first step is to examine developments in each of these 9 industries. Let us focus on those for which data on concentration can be considered reliable. In the case of motor vehicles, he notes a rather steady secular rise, with temporary ups and downs, in the dominance of General Motors, being achieved since 1921 largely at the expense of Ford's share of the market. The gains of General Motors are, he says, not to be attributed to mergers because "the emergence of General Motors Corp. as a leader in the industry came many years after the consolidating operations from 1911 to 1920 under W. C. Durant" [7, p. 36]. This is a strange conclusion in several respects. First, Durant was in and out of control over General Motors in the period from 1908 through 1920; from 1910 to 1915, while out, he built up the Chevrolet Motor Company into a threatening rival [6, pp. 419–429]. After the two were merged in 1915, through financial manipulations by Durant, General Motors' share of the market rose substantially [6, p. 27]. Second, the spectacular rise of General Motors occurred in the immediately following decade of the twenties, after the serious financial problems inherited from Durant's regime had been solved [6, p. 27]. This development cannot be read from the information Weston presents, because he does not give data for the period between 1921 and 1937. Third, merger activity was by no means stopped after 1920 [6, pp. 428–29].

Another important development hidden by Weston's presentation of data is the rapid rise of Chrysler in the late twenties and throughout the thirties, a rise attributable in large measure to mergers. Chrysler's share of the market rose steadily (except for around 1935) from 3 per cent in 1925 to 23 per cent in 1937 [6, p. 27].

For the steel industry he notes a steady decline in the combined share of the four leading producers from 1901 through 1920, an increase through 1930, and a slight decline thereafter. The rise in the twenties he attributes to growth by merger. He says that "since 1930, however, despite continued merger activities, market occupancy of the largest four has decreased slightly" [7, p. 38]. This statement is misleading for two reasons. First, the decline in the combined share of the four (or five) largest firms is barely perceptible, running at 0.2 of a percentage point, well within the range of computational error alone. Sec-

ond, the shares of Republic and Bethlehem, taken individually, increased by a significant amount; and they were the firms with greatest proportional growth by merger over this period. Their combined increase was matched by the decline for U. S. Steel, which had virtually no merger activity in this period.

He describes the trend in the cigarette industry as similar to that in steel. In this case, however, the pattern seems to be one of a general secular decline in the share of each of the four leaders through 1939, with the rise of Philip Morris in the thirties complicating the picture. If the data are carried through 1949 (which is not done by Weston), the pattern is changed to the extent that the largest producer, American Tobacco, regains most of the share it lost between 1912 and 1939 [5, p. 94].

The two remaining industries with reliable data on concentration are aluminum and cement. Weston rightly describes aluminum as a special case of decreasing concentration resulting from disposal of wartime-created capacity. We are given few data on concentration in the cement industry, covering only the span from 1929 through 1945. We are given even fewer data on absolute growth by merger.[14] Hence no conclusions can be drawn about this industry, and Weston does not draw any.

According to Weston, the data for these industries, and the other four not discussed here, "suggest that concentration in industries has not generally been increased by mergers since 1904. In the majority of industries for which information is available, decreases in concentration have actually occurred since 1904 despite merger activity" [7, p. 42]. In a sense Weston is certainly right: the share of the largest firm in most industries has declined, and the combined share of the two, three, or four largest firms has also. But this is not really relevant; the important question is whether the declines occurred for firms with high or low proportional growth by merger. The following table shows a rather consistent relation between declines and relatively low proportional growth by merger, and between increases and relatively high proportional growth by merger. This conclusion is not vitiated if industries with questionable data are included. For every industry except rubber tires and possibly meat packing, relatively low proportional growth by merger is associated with declines in share of the market. For every industry except possibly meat packing and cigarettes (one firm only), relatively high proportional growth by merger is associated with increases in share of the market.

Weston's conclusion—namely, that concentration has decreased

[14] See n. 10 above.

RELATION OF MERGERS TO CHANGES IN INDUSTRIAL CONCENTRATION, 1904–1948

Industry and Firm	Proportional Growth by Merger[a] (%)	Change in % Share of Output[b]	Share of Output, 1948[b] (%)
Motor vehicles			
Chrysler	35	+	22
General Motors	20	+	44
Ford	5	−	20
Steel			
Republic	77	+	9
Jones and Laughlin	56	?	5
Bethlehem	47	+	15
U. S. Steel	10 (75)	−	33
Cigarettes[c]			
American Tobacco	13 (34)	−	31
Philip Morris	12	+	9
Lorillard	6 (58)	−	5
Reynolds	3	−	26
Liggett and Myers	0.1 (16)	−	20
Aluminum			
Reynolds	14	+	33
Permanente	9	+	17
Alcoa	7 (7)	−	50
Electrical machinery			
Westinghouse	25	+?	15?
General Electric	14	−?	25?
Meat packing[d]			
Armour	60	0?	12?
Wilson	35	0?	4?
Swift	13	−?	13?
Cudahy	11	0?	3?
Ribber tires[e]			
Goodrich	85	+?	14?
U. S. Rubber	47 (55)	0?	21?
Goodyear	6	+?	25?
Firestone	5	+?	24?
Tin cans[d]			
Continental Can	47	+?	25?
American Can	4 (34)	−?	51?

[a] Initial assets counted as growth by merger, except for firms engaging in important mergers before 1904 [7, p. 150] and firms in the cigarette industry formed by dissolution of the Tobacco Trust in 1911. For the latter two groups of firms, percentages including initial assets as growth by merger are shown in parentheses. Source: [7, appendix E].

[b] Source: [7, pp. 39–41]; for the cigarette industry, also [5, p. 94].

[c] Terminal date is 1949.

[d] Terminal date is 1939.

[e] Terminal date is 1947.

since 1904 despite mergers—leads him to a second line of inquiry. He raises the possibility that "decreases in concentration might have been even larger in the absence of mergers" [7, p. 44]. He embarks at this point on a statistical analysis that I am not sure I fully understand. My explanation must be given with the warning that it may not be an accurate representation of what Weston is trying to do.

We may get at his approach by considering how information on proportional growth by merger might be married to information on output concentration. Let us suppose that, in the absence of mergers, the growth of a firm would have been smaller by exactly the amount of assets acquired by mergers. Let us further suppose that a firm's output grows in the same percentage as its assets; that is, a doubling of assets leads to a doubling of output. Finally, let us suppose that the firms acquired by merger had retained their separate identities and had not grown. These are heroic assumptions, subject to all kinds of qualification; but they can perhaps serve as working hypotheses for deriving a first approximation, which will almost surely be an underestimate, of the amount of concentration attributable to mergers. If they are accepted on this basis, it follows that the fraction of a firm's share of output attributable to mergers is measured by the fraction of its growth directly accounted for by mergers. For instance, by this reasoning 77 percent of Republic's share of steel output would be attributable to mergers, since 77 per cent of its growth is directly accounted for by mergers; instead of producing 9 per cent of steel output in 1948, it would have produced only 2 per cent if none of the mergers had taken place.

In principle the effect of mergers over any desired period could be estimated in this way, by measuring proportional growth by merger over that period alone. We cannot do this with the statistics worked up by Weston, however; for his measures of proportional growth by merger differ only in their treatments of assets in the initial years for which he found data, years that vary widely among firms [7, p. 11]. By extensive reworking of his basic data we might be able to eliminate all mergers before some particular date, but this would be a major job. Except for that possibility, the only thing that can be done is to estimate the effect of mergers over the entire life of firms. To do this one needs to look no further than the measures of proportional growth by merger, with initial assets counted as growth by merger; that is, all one needs is the evidence Weston has developed on the importance of mergers in absolute growth. Since mergers have generally accounted for substantial fractions of the absolute growth of the firms studied, it follows that they also account for substantial fractions of the firms'

shares of output. Hence they have caused the pattern of concentration to be significantly different from what it otherwise would have been.

Weston looks at the problem somewhat differently. He tries to relate *quantitative changes* in concentration with growth by merger, each occurring over restricted time periods; in addition, he tries to measure the fraction of a firm's share of output around 1948 that is attributable to mergers occurring over a restricted time period. It is hard to see what his findings can be taken to mean. First, the period of mergers will only rarely coincide with the period of changes in concentration, since the initial date in each case is simply the earliest year for which pertinent data could be found, in the one case on assets, in the other on share of output. Second, periods will vary widely among firms. Third, the estimates of shares of output do not have the accuracy required by the analysis. Finally, his sample has dwindled to 25 firms, whose identities are not revealed.

There would seem to be little reason for reviewing his findings, but it may be appropriate to say a few more words about his general method. The assumptions underlying his approach seem to be those outlined above, though he does not state them explicitly, and they are not easy to unravel from his explanation of procedures. Perhaps it is best to let the reader decide by having Weston speak for himself:

> Of the several possible techniques for measuring the influence of internal expansion on existing levels of concentration, the following appeared to be most useful. First, the market occupancy percentage of the leading firms in an industry for the earliest year for which data are available was calculated. Second, data on external growth were deducted from absolute amounts of output or total assets of individual firms, but not for the industry as a whole. Third, the adjusted data of output or total assets for the firms were used to calculate market occupany percentages of individual firms which would have obtained if the growth of the firms had occurred entirely by internal expansion. Fourth, the adjusted concentration ratios were compared with the concentration ratios of the earliest period to measure the extent of present concentration due to internal growth . . . [7, p. 44].
>
> . . . concentration ratios which would have existed in the absence of external growth subsequent to the initial year for which data could be secured for individual firms . . . are [next] deducted from concentration ratios existing in 1947, to provide measures of the extent to which present concentration is accounted for by acquisitions which took place after the early merger movement. [7, p. 46.]

This is the entire explanation. There is no further clarification of how time periods vary; how proportional growth by merger is measured; how assets acquired by merger were "deducted from absolute amounts of output"; how data on concentration of assets were com-

piled, and in what cases they were used; what firms and industries are covered in the sample; and so on.

One must regretfully conclude that Weston's discussion of the role played by mergers in the development of recent patterns of industrial concentration adds little to the knowledge already implicit in his evidence on the importance of mergers in absolute growth. Some new light is shed on developments in three industries, but these make up only a tiny sample of all industries, and much was known about them before. He does not provide the evidence needed to back up his contention that mergers have had very little to do with changes in industrial concentration since 1904.

WESTON *versus* STIGLER

An interesting sidelight to Weston's book is a running argument with Stigler over the relation between mergers and industrial concentration.[15] It deserves mention because of its strong influence on the course of Weston's enquiry.

The primary source of dispute is a paper by Stigler [4]. Weston sees in this paper several controversial theses, which he puts as follows: (*a*) "mergers have been the major factor causing the development of bigness and concentration in the American economy" [7, p. 7]; (*b*) mergers around the turn of the century were motivated solely by a desire to achieve monopoly [7, p. 32]; (*c*) "mergers have been the main instruments by which partial monopolies have been transformed into oligopolies" [7, p. 36]; and (*d*) survival is the "test of relative efficiency among firms of different sizes" [7, pp. 64–65].

Now exegesis is a tricky business, and often a fruitless one; there is little to be gained here by prolonged laboring over what Stigler "really" said. Allow me, however, to offer some brief interpretations of Stigler's views on these points, and to add a few comments on some of Weston's replies. For the rest, let Stigler's paper speak for itself.

Stigler makes a remark in his introductory comments that is akin to the first thesis attributed to him. It is not repeated at any other point, in either the same or related form. Moreover, close examination reveals that the kinship is remote. This is what he says: "There are no large American companies that have not grown somewhat by merger, and

[15] There is another argument as well, which will not be discussed here. It deals with the importance of taking account of the international scope of markets in calculating concentration ratios. Stigler's general position is that failure to do so makes ratios significantly higher than they should be [3, p. 7; also apparently in a letter to Weston]. Weston recomputes ratios for 1935 to include imports in total output and concludes that the ratios generally are not significantly reduced. It is difficult to comment on this dispute because part of it stems from an unpublished letter sent by Stigler to Weston, whose contents are alluded to by Weston.

probably very few that have grown much by the alternative method of internal expansion"; to which is added the footnote: "Unless otherwise indicated, size of the firm is to be measured relative to the size of the industry" [4, p. 23]. I take this to mean that dominant firms would not generally have gotten that way in the absence of mergers. Weston's interpretation, which need not be repeated here, is quite different.

Hard searching has not produced for me the second and third theses. The closest Stigler comes to them is the following statement, which is far away indeed: "We shall find it useful to divide this history [of mergers] into two periods, in which monopoly and oligopoly, respectively, were the primary goals" [4, p. 27].

The last thesis is really there. His full statement is as follows:

> The comparative private costs of firms of various sizes can be measured in only one way: by ascertaining whether firms of the various sizes are able to survive in the industry. Survival is the only test of a firm's ability to cope with all the problems: buying inputs, soothing laborers, finding customers, introducing new products and techniques, coping with fluctuations, evading regulations, etc. A cross-sectional study of the costs of inputs per unit of output in a given period measures only one facet of the firm's efficiency and yields no conclusion on efficiency in the large. Conversely, if a firm of a given size survives, we may infer that its costs are equal to those of other sizes of firm, being neither less (or firms of this size would grow in number relative to the industry) nor more (or firms of this size would decline in number relative to the industry) [4, p. 26].

Weston examines this statement in his chapter on the theory of mergers, which we have not discussed. He disagrees on four grounds. First, firms classified in the same industry, as defined for instance by the Census, do not all produce the same products; in particular, large ones typically produce many products, while small ones typically specialize in a single product, frequently an accessory or a custom item. This is certainly true, and it is a proper warning against incautious use of data. But it does not contradict Stigler's proposition. It might be added parenthetically that, even by taking full advantage of such empirical ambiguities, one is hard-pressed to name more than a handful of large industries in which there are not firms of widely varying size producing essentially the same product for essentially the same markets. Those who are sceptical should try.

Second, small firms may be kept alive because dominant firms spread a price "umbrella" over them; that is, the difference between cost and price for the dominant firm is sufficiently large to allow survival of less efficient small firms. This point may be valid only if smaller

firms account for a very small portion of output. Larger firms, if more efficient, should gradually displace smaller ones, the point made by Stigler. If they do not, the implication is that the cost advantage to the dominant firms, if any, must be unique (in the form of an economic rent), not available to large firms in general.

Third, smaller firms might be satisfied with lower rates of return on investment. This is possible; but so might larger firms. In any event, does this make any difference in the general results? The fourth point is essentially a repetition of this one, suggesting in addition that smaller firms might be satisfied to have their assets undervalued.

The most interesting thing about this controversy is the way in which Weston has misread Stigler, particularly on the first three points. In the paper under dispute, Stigler is really trying to find out, in part by recourse to history, why mergers have been so often used in preference to internal expansion as a means of achieving dominance in an industry; and why they have, since about 1904, contributed more toward oligopoly than toward monopoly. In answering these questions Stigler is led to a general theoretical explanation for industrial concentration, deflating the importance of economies of scale and inflating the importance of temporary exploitation. Here are revealed some of the basic issues that motivate us to seek more data on mergers. If the data are to be relevant, the empirical questions must also be relevant. Somehow, by bare margins in some cases, Weston has failed to raise the relevant questions.

CONCLUDING REMARKS

This review, like most, has stressed a book's vices and slighted its virtues. A few words are called for to help redress the imbalance.

The book has resulted from a major research undertaking. It offers much information not before available, and specialists in the field of industrial organization will surely want to exploit it fully. They will also want to make use of the likely rich source of data represented by Weston's worksheets, obtainable from the Bureau of Business and Economic Research, University of California.

At the same time, the reader must beware of blind acceptance of Weston's conclusions and some of his analysis. The facts as I see them do not support much of what Weston has to say, particularly about the influence of mergers on industrial concentration since 1904. In cases where they do, the conclusions sometimes have a significance different from what Weston supposes.

In brief, this book breaks the ground well in an area where little

comprehensive statistical work had been previously done. It is not the final word, but it is a welcome beginning.

REFERENCES

[1] Edwards, Corwin D., *Maintaining Competition*, New York, McGraw Hill, 1949.

[2] Nutter, G. Warren, *The Extent of Enterprise Monopoly in the United States, 1899–1939*, Chicago, University of Chicago Press, 1951.

[3] Stigler, George J., "The Extent and Bases of Monopoly," *American Economic Review*, XXXII, Supplement (1942), 1–22.

[4] ———, "Monopoly and Oligopoly by Merger," *American Economic Review*, Proceedings, XL (1950), 23–34.

[5] Tennant, Richard B., *The American Cigarette Industry*, New Haven, Yale University Press, 1950.

[6] U. S. Federal Trade Commission, *Report on Motor Vehicle Industry*, Washington, Government Printing Office, 1939.

[7] Weston, J. Fred, *The Role of Mergers in the Growth of Large Firms*, Berkeley and Los Angeles, University of California Press, 1953.

SPURIOUS CORRELATION: A CAUSAL INTERPRETATION*

HERBERT A. SIMON
Carnegie Institute of Technology

To test whether a correlation between two variables is genuine or spurious, additional variables and equations must be introduced, and sufficient assumptions must be made to identify the parameters of this wider system. If the two original variables are causally related in the wider system, the correlation is "genuine."

EVEN in the first course in statistics, the slogan "Correlation is no proof of causation!" is imprinted firmly in the mind of the aspiring statistician or social scientist. It is possible that he leaves the course (and many subsequent courses) with no very clear ideas as to what *is* proved by correlation, but he never ceases to be on guard against "spurious" correlation, that master of imposture who is always representing himself as "true" correlation.

The very distinction between "true" and "spurious" correlation appears to imply that while correlation in general may be no proof of causation, "true" correlation does constitute such proof. If this is what is intended by the adjective "true," are there any operational means for distinguishing between true correlations, which do imply causation, and spurious correlations, which do not?

A generation or more ago, the concept of spurious correlation was examined by a number of statisticians, and in particular by G. U. Yule [8]. More recently important contributions to our understanding of the phenomenon have been made by Hans Zeisel [9] and by Patricia L. Kendall and Paul F. Lazarsfeld [1]. Essentially, all these treatments deal with the three variable case—the clarification of the relation between two variables by the introduction of a third. Generalizations to n variables are indicated but not examined in detail.

Meanwhile, the main stream of statistical research has been diverted into somewhat different (but closely related) directions by Frisch's work on confluence analysis and the subsequent exploration of the "identification problem" and of "structural relations" at the hands of Haavelmo, Hurwicz, Koopmans, Marschak, and many others.[1] This work has been carried on at a level of great generality. It has now reached a point where it can be used to illuminate the concept of

* I am indebted to Richard M. Cyert, Paul F. Lazarsfeld, Roy Radner, and T. C. Koopmans for valuable comments on earlier drafts of this paper.

[1] See Koopmans (2) for a survey and references to the literature.

spurious correlation in the three-variable case. The bridge from the identification problem to the problem of spurious correlation is built by constructing a precise and operationally meaningful definition of causality—or, more specifically, of causal ordering among variables in a model.[2]

1. STATEMENT OF THE PROBLEM

How do we ordinarily make causal inferences from data on correlations? We begin with a set of observations of a pair of variables, x and y. We compute the coefficient of correlation, r_{xy}, between the variables and whenever this coefficient is significantly different from zero we wish to know what we can conclude as to the causal relation between the two variables. If we are suspicious that the observed correlation may derive from "spurious" causes, we introduce a third variable, z, that, we conjecture, may account for this observed correlation. We next compute the partial correlation, $r_{xy\cdot z}$, between x and y with z "held constant," and compare this with the zero order correlation, r_{xy}. If $r_{xy\cdot z}$ is close to zero, while r_{xy} is not, we conclude that either: (a) z is an intervening variable—the causal effect of x on y (or vice versa) operates through z; or (b) the correlation between x and y results from the joint causal effect of z on both those variables, and hence this correlation is spurious. It will be noted that in case (a) we do not know whether the causal arrow should run from x to y or from y to x (via z in both cases); and in any event, the correlations do not tell us whether we have case (a) or case (b).

The problem may be clarified by a pair of specific examples adapted from Zeisel.[3]

I. The data consist of measurements of three variables in a number of groups of people: x is the percentage of members of the group that is married, y is the average number of pounds of candy consumed per month per member, z is the average age of members of the group. A high (negative) correlation, r_{xy}, was observed between marital status

[2] Simon (6) and (7). See also Orcutt (4) and (5). I should like, without elaborating it here, to insert the *caveat* that the concept of causal ordering employed in this paper does not in any way solve the "problem of Hume" nor contradict his assertion that all we can ever observe are covariations. If we employ an ontological definition of cause—one based on the notion of the "necessary" connection of events—then correlation cannot, of course, prove causation. But neither can anything else prove causation, and hence we can have no basis for distinguishing "true" from "spurious" correlation. If we wish to retain the latter distinction (and working scientists have not shown that they are able to get along without it), and if at the same time we wish to remain empiricists, then the term "cause" must be defined in a way that does not entail objectionable ontological consequences. That is the course we shall pursue here.

[3] Zeisel [9], pp. 192–95. Reference to the original source will show that in this and the following example we have changed the variables from attributes to continuous variables for purposes of exposition.

and amount of candy consumed. But there was also a high (negative) correlation, r_{yz}, between candy consumption and age; and a high (positive) correlation, r_{xz}, between marital status and age. However, when age was held constant, the correlation $r_{xy \cdot z}$, between marital status and candy consumption was nearly zero. By our previous analysis, either age is an intervening variable between marital status and candy consumption; or the correlation between marital status and candy consumption is spurious, being a joint effect caused by the variation in age. "Common sense"—the nature of which we will want to examine below in detail—tells us that the latter explanation is the correct one.

II. The data consist again of measurements of three variables in a number of groups of people: x is the percentage of female employees who are married, y is the average number of absences per week per employee, z is the average number of hours of housework performed per week per employee.[4] A high (positive) correlation, r_{xy}, was observed between marriage and absenteeism. However, when the amount of housework, z was held constant, the correlation $r_{xy \cdot z}$ was virtually zero. In this case, by applying again some common sense notions about the direction of causation, we reach the conclusion that z is an intervening variable between x and y: that is, that marriage results in a higher average amount of housework performed, and this, in turn, in more absenteeism.

Now what is bothersome about these two examples is that the same statistical evidence, so far as the coefficients of correlation are concerned, has been used to reach entirely different conclusions in the two cases. In the first case we concluded that the correlation between x and y was spurious; in the second case that there was a true causal relationship, mediated by the intervening variable z. Clearly, it was not the statistical evidence, but the "common sense" assumptions added afterwards, that permitted us to draw these distinct conclusions.

2. CAUSAL RELATIONS

In investigating spurious correlation we are interested in learning whether the relation between two variables persists or disappears when we introduce a third variable. Throughout this paper (as in all ordinary correlation analyses) we will assume that the relations in question are linear, and without loss of generality, that the variables are measured from their respective means.

[4] Zeisel [9], pp. 191–92.

Now suppose we have a system of three variables whose behavior is determined by some set of linear mechanisms. In general we will need three mechanisms, each represented by an equation—three equations to determine the three variables. One such set of mechanisms would be that in which each of the variables *directly influenced* the other two. That is, in one equation x would appear as the dependent variable, y and z as independent variables; in the second equation y would appear as the dependent variable, x and z as the independent variables; in the third equation, z as dependent variable, x and y as independent variables.[5]

The equations would look like this:

$$\text{(I)}\quad \begin{aligned} &(2.1) && x + a_{12}y + a_{13}z = u_1, \\ &(2.2) && a_{21}x + y + a_{23}z = u_2, \\ &(2.3) && a_{31}x + a_{32}y + z = u_3, \end{aligned}$$

where the u's are "error" terms that measure the net effects of all other variables (those not introduced explicitly) upon the system. We refer to $A = \|a_{ij}\|$ as the *coefficient matrix* of the system.

Next, let us suppose that not all the variables directly influence all the others—that some independent variables are absent from some of the equations. This is equivalent to saying that some of the elements of the coefficient matrix are zero. By way of specific example, let us assume that $a_{31} = a_{32} = a_{21} = 0$. Then the equation system (I) reduces to:

$$\text{(II)}\quad \begin{aligned} &(2.4) && x + a_{12}y + a_{13}z = u_1, \\ &(2.5) && y + a_{23}z = u_2, \\ &(2.6) && z = u_3. \end{aligned}$$

By examining the equations (II), we see that a change in u_3 will change the value of z directly, and the values of x and y indirectly; a change in u_2 will change y directly and x indirectly, but will leave z unchanged; a change in u_1 will change only x. Then we may say that y *is causally dependent on* z in (II), and that x is causally dependent on y and z.

If x and y were correlated, we would say that the correlation was genuine in the case of the system (II), for $a_{12} \neq 0$. Suppose, instead, that the system were (III):

[5] The question of how we distinguish between "dependent" and "independent" variables is discussed in Simon (7), and will receive further attention in this paper.

$$\text{(III)}\quad \begin{aligned} &(2.7) & x + a_{13}z &= u_1, \\ &(2.8) & y + a_{23}z &= u_2, \\ &(2.9) & z &= u_3. \end{aligned}$$

In this case we would regard the correlation between x and y as spurious, because it is due solely to the influence of z on the variables x and y. Systems (II) and (III) are, of course, not the only possible cases, and we shall need to consider others later.

3. THE *a priori* ASSUMPTIONS

We shall show that the decision that a partial correlation is or is not spurious (does not or does indicate a causal ordering) can in general only be reached if *a priori* assumptions are made that certain *other* causal relations do *not* hold among the variables. This is the meaning of the "common sense" assumptions mentioned earlier. Let us make this more precise.

Apart from any statistical evidence, we are prepared to assert in the first example of Section 1 that the age of a person does *not* depend upon either his candy consumption or his marital status. Hence z cannot be causally dependent upon either x or y. This is a genuine empirical assumption, since the variable "chronological age" really stands, in these equations, as a surrogate for physiological and sociological age. Nevertheless, it is an assumption that we are quite prepared to make on evidence apart from the statistics presented. Similarly, in the second example of Section 1, we are prepared to assert (on grounds of other empirical knowledge) that marital status is not causally dependent upon either amount of housework or absenteeism.[4]

The need for such *a priori* assumption follows from considerations of elementary algebra. We have seen that whether a correlation is genuine or spurious depends on which of the coefficients, a_{ij}, of A are zero, and which are non-zero. But these coefficients are not observable nor are the "error" terms, u_1, u_2 and u_3. What we observe is a sample of values of x, y, and z.

Hence, from the standpoint of the problem of statistical estimation, we must regard the $3n$ sample values of x, y, and z as numbers given by observation, and the $3n$ error terms, u_i, together with the six coefficients, a_{ij}, as variables to be estimated. But then we have $(3n+6)$

[4] Since these are empirical assumptions it is conceivable that they are wrong, and indeed, we can imagine mechanisms that would reverse the causal ordering in the second example. What is argued here is that these assumptions, right or wrong, are implicit in the determination of whether the correlation is true or spurious.

variables ($3n$ u's and six a's) and only $3n$ equations (three for each sample point). Speaking roughly in "equation-counting" terms, we need six more equations, and we depend on the *a priori* assumptions to provide these additional relations.

The *a priori* assumptions we commonly employ are of two kinds:

(1) *A priori* assumptions that certain variables are not directly dependent on certain others. Sometimes such assumptions come from knowledge of the time sequence of events. That is, we make the general assumption about the world that if y precedes x in time, then $a_{21}=0$—x does not directly influence y.

(2) *A priori* assumptions that the errors are uncorrelated—i.e., that "all other" variables influencing x are uncorrelated with "all other" variables influencing y, and so on. Writing $E(u_iu_j)$ for the expected value of u_iu_j, this gives us the three additional equations:

$$E(u_1u_2) = 0; \qquad E(u_1u_3) = 0; \qquad E(u_2u_3) = 0.$$

Again it must be emphasized that these assumptions are "a priori" only in the sense that they are not derived from the statistical data from which the correlations among x, y, and z are computed. The assumptions are clearly empirical.

As a matter of fact, it is precisely because we are unwilling to make the analogous empirical assumptions in the two-variable case (the correlation between x and y alone) that the problem of spurious correlation arises at all. For consider the two-variable system:

$$\text{(IV)}\quad \begin{array}{ll} (3.1) & x + b_{12}y = v_1 \\ (3.2) & y = v_2 \end{array}$$

We suppose that y precedes x in time, so that we are willing to set $b_{21}=0$ by an assumption of type (1). Then, if we make the type (2) assumption that $E(v_1v_2)=0$, we can immediately obtain a unique estimate of b_{12}. For multiplying the two equations, and taking expected values, we get:

$$(3.3)\qquad E(xy) + b_{12}E(y^2) = E(v_1v_2) = 0.$$

Whence

$$(3.4)\qquad b_{12} = -\frac{E(xy)}{E(y^2)} = -\frac{\sigma_y}{\sigma_x}r_{xy}.$$

It follows immediately that (sampling questions aside) b_{12} will be zero or non-zero as r_{12} is zero or non-zero. *Hence correlation is proof of causa-*

tion in the two-variable case if we are willing to make the assumptions of time precedence and non-correlation of the error terms.

If we suspect the correlation to be spurious, we look for a common component, z, of v_1 and v_2 which might account for their correlation:

$$v_1 \equiv u_1 - a_{13}z, \tag{3.5a}$$

$$v_2 \equiv u_2 - a_{23}z. \tag{3.5b}$$

Substitution of these relations in (IV) brings us back immediately to systems like (II). This substitution replaces the unobservable v's by unobservable u's. Hence, we are not relieved of the necessity of postulating independence of the errors. We are more willing to make these assumptions in the three-variable case because we have explicitly removed from the error term the component z which we suspect is the source, if any, of the correlation of the v's.

Stated otherwise, introduction of the third variable, z, to test the genuineness or spuriousness of the correlation between x and y, is a method for determining whether in fact the v's of the original two variable system were uncorrelated. But the test can be carried out only on the assumption that the unobservable error terms of the three variable system are uncorrelated. If we suspect this to be false, we must further enlarge the system by introduction of a fourth variable, and so on, until we obtain a system we are willing to regard as "complete" in this sense.

Summarizing our analysis we conclude that:

(1) Our task is to determine which of the six off-diagonal matrix coefficients in a system like (I) are zero.

(2) But we are confronted with a system containing a total of nine variables (six coefficients and three unobservable errors), and only three equations.

(3) Hence we must obtain six more relations by making certain *a priori* assumptions.

(a) Three of these relations may be obtained, from considerations of time precedence of variables or analogous evidence, in the form of direct assumptions that three of the a_{ij} are zero.

(b) Three more relations may be obtained by assuming the errors to be uncorrelated.

4. SPURIOUS CORRELATION

Before proceeding with the algebra, it may be helpful to look a little more closely at the matrix of coefficients in systems like (I), (II), and

(III), disregarding the numerical values of the coefficients, but considering only whether they are non-vanishing (X), or vanishing (0). An example of such a matrix would be

$$\begin{Vmatrix} X & 0 & 0 \\ X & X & X \\ 0 & 0 & X \end{Vmatrix}.$$

In this case x and z both influence y, but not each other, and y influences neither x nor z. Moreover, a change in u_2—u_1 and u_3 being constant—will change y, but not x or z; a change in u_1 will change x and y, but not z; a change in u_3 will change z and y, but not x. Hence the causal ordering may be depicted thus:

$$\begin{tikzcd} x \arrow[dr] & & z \arrow[dl] \\ & y & \end{tikzcd}$$

In this case the correlation between x and y is "true," and not spurious.

Since there are six off-diagonal elements in the matrix, there are $2^6=64$ possible configurations of X's and 0's. The *a priori* assumptions (1), however, require 0's in three specified cells, and hence for each such set of assumptions there are only $2^3=8$ possible distinct configurations. If (to make a definite assumption) x does not depend on y, then there are three possible orderings of the variables (z, x, y; x, z, y; x, y, z), and consequently $3\cdot 8=24$ possible configurations, but these 24 configurations are not all distinct. For example, the one depicted above is consistent with either the ordering (z, x, y) or the ordering (x, z, y).

Still assuming that x does not depend on y, we will be interested, in particular, in the following configurations:

$$\begin{Vmatrix} X & 0 & 0 \\ X & X & X \\ 0 & 0 & X \end{Vmatrix} \qquad \begin{Vmatrix} X & 0 & X \\ X & X & 0 \\ 0 & 0 & X \end{Vmatrix} \qquad \begin{Vmatrix} X & 0 & 0 \\ X & X & 0 \\ X & 0 & X \end{Vmatrix}$$

(α) (β) (γ)

$$\begin{Vmatrix} X & 0 & X \\ 0 & X & X \\ 0 & 0 & X \end{Vmatrix} \qquad \begin{Vmatrix} X & 0 & 0 \\ 0 & X & X \\ X & 0 & X \end{Vmatrix}$$

(δ) (ϵ)

In Case α, either x may precede z, or z, x. In Cases β and δ, z precedes x; in Cases γ and ϵ, x precedes z. The causal orderings that may be inferred are:

$$\begin{tikzcd} x \arrow[dr] & & z \arrow[dl] \\ & y & \end{tikzcd} \qquad \begin{tikzcd} z \arrow[d] \\ x \arrow[d] \\ y \end{tikzcd} \qquad \begin{tikzcd} & x \arrow[dl] \arrow[dr] & \\ y & & z \end{tikzcd} \qquad \begin{tikzcd} & z \arrow[dl] \arrow[dr] & \\ x & & y \end{tikzcd} \qquad \begin{tikzcd} x \arrow[d] \\ z \arrow[d] \\ y \end{tikzcd}$$

(α) (β) (γ) (δ) (ε)

The two cases we were confronted with in our earlier examples of Section 1 were δ and ϵ, respectively. Hence, δ is the case of spurious correlation due to z; ϵ the case of true correlation with z as an intervening variable.

We come now to the question of which of the matrices that are consistent with the assumed time precedence is the correct one. Suppose, for definiteness, that z precedes x, and x precedes y. Then $a_{12} = a_{31} = a_{32} = 0$; and the system (I) reduces to:

$$x + a_{13}z = u_1, \tag{4.1}$$

$$a_{21}x + y + a_{23}z = u_2, \tag{4.2}$$

$$z = u_3. \tag{4.3}$$

Next, we assume the errors to be uncorrelated:

$$E(u_1u_2) = E(u_1u_3) = E(u_2u_3) = 0. \tag{4.4}$$

Multiplying equations (4.1)–(4.3) by pairs, and taking expected values we get:

$$a_{21}E(x^2) + E(xy) + a_{23}E(xz) + a_{13}[a_{21}E(xz) + E(yz) + a_{23}E(z^2)] = E(u_1u_2) = 0, \tag{4.5}$$

$$E(xz) + a_{13}E(z^2) = E(u_1u_3) = 0, \tag{4.6}$$

$$a_{21}E(xz) + E(yz) + a_{23}E(z^2) = E(u_2u_3) = 0. \tag{4.7}$$

Because of (4.7), the terms in the bracket of (4.5) vanish, giving:

$$a_{21}E(x^2) + E(xy) + a_{23}E(xz) = 0. \tag{4.8}$$

Solving for $E(xz)$, $E(yz)$ and $E(xy)$ we find:

$$E(xz) = -a_{13}E(z^2), \tag{4.9}$$

$$E(yz) = (a_{13}a_{21} - a_{23})E(z^2), \tag{4.10}$$

$$E(xy) = a_{13}a_{23}E(z^2) - a_{21}E(x^2). \tag{4.11}$$

Case α: Now in the matrix of case α, above, we have $a_{13}=0$. Hence:

(4.12a) $$E(xz) = 0; \text{(4.12b)} E(yz) = -a_{23}E(z^2),$$

(4.12c) $$E(xy) = -a_{21}E(x^2).$$

Case β: In this case, $a_{23}=0$, hence,

(4.13a) $$E(xz) = -a_{13}E(z^2); \text{(4.13b)} E(yz) = a_{13}a_{21}E(z^2);$$

(4.13c) $$E(xy) = -a_{21}E(x^2);$$

from which it also follows that:

(4.14) $$E(xy) = E(x^2)\frac{E(yz)}{E(xz)}.$$

Case δ: In this case, $a_{21}=0$. Hence,

(4.15a) $$E(xz) = -a_{13}E(z^2); \text{(4.15b)} E(yz) = -a_{23}E(z^2);$$

(4.15c) $$E(xy) = a_{13}a_{23}E(z^2);$$

and we deduce also that:

(4.16) $$E(xy) = \frac{E(xz)E(yz)}{E(z^2)}.$$

We have now proved that $a_{13}=0$ implies (4.12a); that $a_{23}=0$ implies (4.14); and that $a_{21}=0$ implies (4.16). We shall show that the converse also holds.

To prove that (4.12a) implies $a_{13}=0$ we need only set the left-hand side of (4.9) equal to zero.

To prove that (4.14) implies that $a_{23}=0$ we substitute in (4.14) the values of the cross-products from (4.9)–(4.11). After some simplification, we obtain:

(4.17) $$a_{23}[E(x^2) - a_{13}^2E(z^2)] = 0.$$

Now since, from (4.1)

(4.18) $$E(x^2) - E(u_1^2) + 2a_{13}E(zu_1) = a_{13}^2E(z^2),$$

and since, by multiplying (4.3) by u_1, we can show that $E(zu_1)=0$, the second factor of (4.17) can vanish only in case $E(u_1^2)=0$. Excluding this degenerate case, we conclude that $a_{23}=0$.

To prove that (4.16) implies that $a_{21}=0$, we proceed in a similar manner, obtaining:

(4.19) $$a_{21}[E(x^2) - a_{13}^2E(z^2)] = 0,$$

from which we can conclude that $a_{21}=0$.

We can summarize the results as follows:

1) If $E(xz)=0$, $E(yz)\neq 0$, $E(xy)\neq 0$, we have Case α

2) If none of the cross-products is zero, and

$$E(xy) = E(x^2)\frac{E(yz)}{E(xz)},$$

we have Case β.

3) If none of the cross-products is zero, and

$$E(xy) = \frac{E(xz)E(yz)}{E(z^2)},$$

we have Case δ.

We can combine these conditions to find the conditions that two or more of the coefficients a_{13}, a_{23}, a_{21} vanish:

4) If $a_{13}=a_{23}=0$, we find that:
$E(xz)=0$, $E(yz)=0$. Call this Case $(\alpha\beta)$.

5) If $a_{13}=a_{21}=0$, we find that:
$E(xz)=0$, $E(xy)=0$. Call this Case $(\alpha\delta)$.

6) If $a_{23}=a_{21}=0$, we find that:
$E(yz)=0$, $E(xy)=0$. Call this Case $(\beta\delta)$.

7) If $a_{13}=a_{23}=a_{21}=0$, then
$E(xz)=E(yz)=E(xy)=0$. Call this Case $(\alpha\beta\delta)$.

8) If none of the conditions (1)–(7) are satisfied, then all three coefficients a_{13}, a_{23}, a_{21} are non-zero. Thus, by observing which of the conditions (1) through (8) are satisfied by the expected values of the cross products, we can determine what the causal ordering is of the variables.[7]

We can see also, from this analysis, why the vanishing of the partial correlation of x and y is evidence for the spuriousness of the zero-order correlation between x and y. For the numerator of the partial correlation coefficient $r_{xy\cdot z}$, we have:

$$N(r_{xy\cdot z}) = \frac{E(xy)}{\sqrt{E(x^2)E(y^2)}} - \frac{E(xz)E(yz)}{E(z^2)\sqrt{E(x^2)E(y^2)}}. \tag{4.20}$$

We see that the condition for Case δ is precisely that $r_{xy\cdot z}$ vanish while none of the coefficients, r_{xy}, r_{xz}, r_{yz} vanish. From this we conclude

[7] Of course, the expected values are not, strictly speaking, observables except in a probability sense. However, we do not wish to go into sampling questions here, and simply assume that we have good estimates of the expected values.

that the first illustrative example of Section 1 falls in Case δ, as previously asserted. A similar analysis shows that the second illustrative example of Section 1 falls in Case ϵ.

In summary, our procedure for interpreting, by introduction of an additional variable z, the correlation between x and y consists in making the six *a priori* assumptions described earlier; estimating the expected values, $E(xy)$, $E(xz)$, and $E(yz)$; and determining from their values which of the eight enumerated cases holds. Each case corresponds to a specified arrangement of zero and non-zero elements in the coefficient matrix and hence to a definite causal ordering of the variables.

5. THE CASE OF EXPERIMENTATION

In sections (3)–(4) we have treated u_1, u_2 and u_3 as random variables. The causal ordering among x, y, and z can also be determined without *a priori* assumptions in the case where u_1, u_2, and u_3 are controlled by an experimenter. For simplicity of illustration we assume there is time precedence among the variables. Then the matrix is triangular, so that $a_{ij}\neq 0$ implies $a_{ji}=0$; and $a_{ij}\neq 0$, $a_{jk}\neq 0$ implies $a_{ki}=0$.

Under the given assumptions at least three of the off-diagonal a's in (I) must vanish, and the equations and variables can be reordered so that all the non-vanishing coefficients lie on or below the diagonal. If (with this ordering) u_2 or u_3 are varied, at least the variable determined by the first equation will remain constant (since it depends only on u_1). Similarly, if u_3 is varied, the variables determined by the first and second equations will remain constant.

In this way we discover which variables are determined by which equations. Further, if varying u_i causes a particular variable other than the ith to change in value, this variable must be causally dependent on the ith.

Suppose, for example, that variation in u_1 brings about a change in x and y, variation in u_2 a change in y, and variation in u_3 a change in x, y, and z. Then we know that y is causally dependent upon x and z, and x upon z. But this is precisely the Case β treated previously under the assumption that the u's were stochastic variables.

6. CONCLUSION

In this paper I have tried to clarify the logical processes and assumptions that are involved in the usual procedures for testing whether a correlation between two variables is true or spurious. These procedures begin by imbedding the relation between the two variables in a larger three-variable system that is assumed to be self-contained,

except for stochastic disturbances or parameters controlled by an experimenter.

Since the coefficients in the three-variable system will not in general be identifiable, and since the determination of the causal ordering implies identifiability, the test for spuriousness of the correlation requires additional assumptions to be made. These assumptions are usually of two kinds. The first, ordinarily made explicit, are assumptions that certain variables do *not* have a causal influence on certain others. These assumptions reduce the number of degrees of freedom of the system of coefficients by implying that three specified coefficients are zero.

The second type of assumption, more often implicit than explicit, is that the random disturbances associated with the three-variable system are uncorrelated. This assumption gives us a sufficient number of additional restrictions to secure the identifiability of the remaining coefficients, and hence to determine the causal ordering of the variables.

REFERENCES

[1] Kendall, Patricia L., and Lazarsfeld, Paul F., "Problems of Survey Analysis," in Merton and Lazarsfeld (eds.), *Continuities in Social Research*, The Free Press, 1950, 133–96.

[2] Koopmans, Tjalling C., "Identification Problems in Economic Model Construction," *Econometrica* 17: 125–44 (April 1949), reprinted as Chapter II in *Studies in Econometric Methods*, Cowles Commission Monograph 14.

[3] ———, "When Is an Equation System Complete for Statistical Purposes?" Chapter 17 in *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10.

[4] Orcutt, Guy H., "Toward Partial Redirection of Econometrics," *The Review of Economics and Statistics*, 34 (1952), 195–213.

[5] ———, "Actions, Consequences, and Causal Relations," *The Review of Economics and Statistics*, 34 (1952), 305–14.

[6] Simon, Herbert A., "On the Definition of the Causal Relation," *The Journal of Philosophy*, 49 (1952), 517–28.

[7] ———, "Causal Ordering and Identifiability." Chapter III in *Studies in Econometric Methods*, Cowles Commission Monograph 14.

[8] Yule, G. Udny, *An Introduction to the Theory of Statistics*, Charles Griffin and Co., 10th ed., 1932, Chapters 4, 12. (Equivalent chapters will be found in all subsequent editions of Yule and Yule and Kendall, through the 14th.)

[9] Zeisel, Hans, *Say It With Figures*, New York, Harper and Brothers, 1947.

EMPIRICAL STUDY OF THE ACCURACY OF SELECTED METHODS OF PROJECTING STATE POPULATIONS*

HELEN R. WHITE

United States Bureau of Agricultural Economics

As tentative guides in the preparation of population projections for geographic subdivisions of the United States, the accuracy in the past of several methods of projecting population has been measured. These measures have been analyzed to some extent for information on the effects of selected factors other than methodology on the accuracy of projections.

ERRORS in particular population projections have been noted and analyzed in the literature of this field, and various methods of projecting population have been evaluated.[1] Nevertheless, little has been written on the accuracy of population projections in general or about the effects of such factors as the size of the base population, past migration rates, and length of the projection period on the accuracy of population projections. The test described below was undertaken in order to provide some guides in deciding what methods should be used in projecting the populations of geographic subdivisions of the United States and in deciding whether any projections should be prepared in certain cases.

Design of test.—The study is based on a comparison of projections to 1940 and 1950 of the 1930 Census population, for each state and for the District of Columbia, prepared by various methods, with the Census data for those dates. Since the projections prepared by Thompson and Whelpton and published in *Estimates of Future Population by States* (National Resources Board, 1934) are based on the 1930 Census, these projections could be used to represent the cohort-survival method.[2] The other methods by which projections have been prepared are limited to those which did not involve extensive computing and

* This paper is based on a project of the Bureau of the Census carried on while the author was employed at that agency. The assistance of Mrs. Beatrice M. Rosen of the Bureau of the Census is gratefully acknowledged. A summary of the results of this study was presented at a meeting of the Population Association of America on April 19, 1952.

[1] See Appendix.

[2] The cohort-survival method involves making separate allowances for changes in each of its age cohorts resulting from mortality and immigration; the initial size of cohorts born after the base date are usually based on projected age-specific or cohort birth rates. More detailed descriptions of specific applications of this method are given in: P. K. Whelpton, "An Empirical Method of Calculating Future Population," *Journal of the American Statistical Association*, 31 (1936), pp. 457–73; P. K. Whelpton, Hope Tisdale Eldridge, and Jacob S. Siegel, *Forecasts of the Population of the United States, 1945–1975*, U. S. Government Printing Office, Washington, 1947; and Jacob S. Siegel and Helen R. White, "Illustrative Projections of the Population of the United States, 1950 to 1960," *Current Population Reports*, Series P-25, No. 43, U. S. Bureau of the Census, August 1950.

for which the results would not be biased by the worker's knowledge of population trends since 1930. These methods are the geometric, arithmetic, apportionment, and ratio methods.

Both the apportionment method and the ratio method require independent projections of the total population of the United States for 1940 and 1950. For this purpose, the sum of the Thompson-Whelpton projections for states, with an allowance for migration, and the actual decennial census national totals[3] were used. Although the cohort-survival, geometric, and arithmetic methods do not necessarily involve the use of independent national totals, the latter two were also adjusted to the Thompson-Whelpton projections and all three were adjusted to the census national totals.

The independent projection of the total population of the United States has been used as a control total (that is, the state projections have been forced to sum to the independent national projection) in the instances mentioned above. Such use is not an inherent characteristic of the ratio method but arises from the adjustment of the appropriate ratios to sum to 1.00. Projections were also obtained by one variation of the ratio method without adjustment of the ratios. These projections are referred to as "Ratio III (unadjusted)" in the text and are presented in the tables under "Unadjusted to national total."

As the preparation of projections is not justified unless the results are better than those obtained by using the available current figures on the size of the population, measures of the errors involved in using the 1930 Census data for 1940 and 1950 have also been developed. The figures are presented in the various tables (under the designation of "Constant") along with the results for the various methods.

Thus the tables present results for the following methods and assumptions:

I. *Unadjusted to national total.*

1. *Cohort-survival, with migration*, the cohort-survival method, assuming continuation of internal migration like 1920–30 (see *Estimates of Future Population by States*, mentioned above, for information on the basic assumptions).

2. *Cohort-survival, no migration*, the same as (1) above except for the assumption of no internal migration.

3. *Geometric* method, assuming the continuation of the 1920–30 average annual rate of increase.

[3] The 1950 Census total was adjusted to include members of the armed forces overseas except those inducted in the Territories and possessions. The 1940 total was not adjusted because of the small number of armed forces involved.

4. *Arithmetic* method, assuming continuation of the 1920–30 average amount of increase per year.

5. *Constant*, the 1930 enumerated population.

6. *Ratio III* (*1900 to 1930 modified*), *T & W national projection*, the ratio method, using the rules presented in *Current Population Reports*, Series P-25, No. 56,[4] for selecting the period used in computing the initial change in the ratio of the population of each division to the total population of the United States and the ratio of the population of each state to the population of its divisions; these rules, which in this case were applied to 1900–30, 1910–30, and 1920–30, first eliminate any period during which the given ratio did not either constantly increase or constantly decrease, and then select, from the remaining periods, the one for which the absolute value of the average annual rate of change in the given ratio was least. It was also assumed that the annual rate of change of each ratio would decrease linearly to zero within fifty years; i.e., by 1980. As these ratios were not adjusted to sum to 1.00, the projected state populations do not sum to the Thompson-Whelpton national projection to which the ratios were applied.

7. *Ratio III* (*1900 to 1930 modified*), *census count*, the same as (6) above except that the projected ratios were applied to the census national totals.

II. *Adjusted to T & W national projection*, using as a control total the independent T &W projection (with internal migration) of the total population of the United States.

1. *Geometric*, the state projections of (I, 3) above adjusted proportionately to sum to the T & W national projection.

2. *Arithmetic*, the state projections of (I, 4) above adjusted proportionately to sum to the T & W national projection.

3. *Apportionment* method, assuming (a) that the increase in the total population of the United States, as indicated by the T & W projection, would be distributed in accordance with the distribution of the 1920–30 increase among those states whose population gained between 1920 and 1930 and assuming (b) that the populations of those states in which there was a decrease during that period would remain constant.

4. *Ratio I*, (*1870 to 1930*), the ratio method, involving the projection of the ratio of the total population of each state to the total population of the United States on the assumptions (a) that the initial change in the ratio would be the same as the 1870–1930 average annual rate of

[4] Helen L. White and Jacob S. Siegel, "Projections of the Population by States: 1955 and 1960." *Current Population Reports*, Series P-25, No. 56, Bureau of the Census, January 1952.

change in the ratio, and (b) that the annual rate of change in the ratio would decrease linearly to zero by 1975; the projected ratios were adjusted to sum to 1.00 and were then applied to the T & W national projection.

5. *Ratio II* (*1930*), the ratio method, assuming that the ratios would remain at the 1930 level; this assumes that the per cent increase in the population of each state would be the same as that of the T & W national projection to which the assumed ratios were applied.

6. *Ratio III* (*1900 to 1930 modified*), using the same assumptions as for (I, 6). The ratios were adjusted to sum to 1.00 and were then applied to the T & W national projection.

III. *Adjusted to census count*, using the census count as a control total.

1. *Cohort-survival, with migration*, the state projections of (I, 1) adjusted proportionately to sum to the census count of the total population.

2. *Cohort-survival, no migration*, the state projections of (I, 2) adjusted proportionately to sum to the census count.

3. *Geometric*, the state projections of (I, 3) adjusted proportionately to sum to the census count.

4. *Arithmetic*, the state projections of (I, 4) adjusted proportionately to sum to the census count.

5. *Apportionment*, assuming (a) that the actual increase in the total population of the United States, from the census count, would be distributed in accordance with the distribution of the 1920–30 increase among those states which gained population during that period and (b) that the population of those states in which there was a decrease during that period would remain constant.

6. *Ratio I* (*1870 to 1930*), applying the ratios of (II, 4) to the census count.

7. *Ratio II* (*1930*), applying the ratios of (II, 5) to the census count.

8. *Ratio III* (*1900 to 1930 modified*), using the same assumptions as for (I, 6). The ratios were adjusted to sum to 1.00 and were then applied to the census count.

As mentioned previously, the projections for 1940 and for 1950, for each state, described above, were compared with the 1940 and 1950 Census returns. Before the comparison was made, the 1950 Census data for each state were adjusted to include members of the armed forces who resided in the given state at the time of induction and to exclude members of the armed forces stationed in the given state who

did not reside there at the time of induction.[5] The deviations of the projections from the census data are summarized in Table 1, which shows the average per cent error (average of the absolute values of the per cent deviations), the maximum per cent error (absolute value), the proportion of errors of ten per cent or more (absolute values), and the proportion of positive errors (over-estimates).

It must be kept in mind that the results presented here are only very rough guides for future periods.

Accuracy of various methods.—The various methods are evaluated on the basis of the summary measures shown in Table 1 for the unadjusted projections and the projections adjusted to the Thompson-Whelpton national projections. The projections adjusted to the census counts are not considered, since census data could not be used for this purpose in actual practice.

One of the rather interesting results of this study is that the cohort-survival method does not appear to yield definitely superior results. (It must be remembered, however, that this method has several advantages over other methods not wholly dependent on its absolute validity.) In fact, no one method is clearly superior to all other methods tested and only the Ratio I method is clearly inferior. The Ratio I method is probably inferior because the basic assumptions place too much emphasis on population change in relatively remote periods; none of the other projections depend so much on population change prior to 1900.

For 1940, the apportionment, the cohort-survival (with migration), the Ratio II, and the Ratio III (unadjusted) results are the best on the basis of average per cent error. On the basis of the proportion of errors of 10 per cent or more, the apportionment, the cohort-survival (with migration), and the Ratio III (both unadjusted and adjusted) results are the best.

For 1950, the arithmetic (unadjusted), the Ratio III (unadjusted),

[5] The effects of military migration during the past decade were removed because they are believed to represent an abnormality which ordinarily could not be taken into account by any of the methods of projecting population. Dr. Henry S. Shryock, Jr., has commented: "At first I thought that your adjusting the Census data to what we sometimes call the *de jure* population level was the wrong thing to do here. The members of the armed forces stationed in the several states are the result of what may be viewed as military migration. Since we are usually interested in forecasting the number that the Census will count at a future date, it would seem appropriate to include the armed forces where they would be enumerated. Furthermore, some members of the armed forces were stationed in the several states in 1930. On the other hand, I realize what you are trying to do is to remove direct effects of the defense preparations on the distribution of population among the states. This attempt is consistent with the usual practice in national projections to assume that a war will not be going on or be in prospect at the future dates for which projections are made. Of course, if the cold war continues long enough, we may come to consider the resulting size and distribution of the armed forces as normal, and in this case we might hesitate to predict the future distribution of population under peaceful conditions."

the cohort-survival (with migration), the apportionment, and the Ratio II results are the best on the basis of average per cent error, each of these methods having an APE under 13. On the basis of the propor-

TABLE 1.—SUMMARY OF PER CENT ERRORS OF PROJECTIONS TO 1940 AND 1950 OF THE POPULATIONS OF THE STATES, FOR SELECTED METHODS

Method	Average per cent error		Maximum per cent error		Proportion of errors of 10 per cent or more (expressed as a per cent)		Proportion of positive errors (expressed as a per cent)	
	1940	1950	1940	1950	1940	1950	1940	1950
Unadjusted to national total								
Cohort-survival (T&W)								
With migration	5.14	12.52	22.64	39.52	12.2	51.0	40.8	14.3
No migration	6.04	15.11	26.41	45.32	20.4	55.1	61.2	34.7
Geometric	8.19	13.31	34.50	47.40	32.7	53.1	69.4	53.1
Arithmetic	6.30	10.91	19.32	38.38	22.4	51.0	67.3	49.0
Constant	8.39	19.27	26.58	46.89	28.6	73.5	12.2	8.2
Ratio III (1900 to 1930 modified)								
T&W national projection*	5.80	12.46	22.72	34.67	14.3	49.0	36.7	16.3
Census count	5.83	9.83	22.83	34.45	14.3	36.7	36.7	44.9
*Adjusted to T&W national projection**								
Geometric	7.60	17.17	25.07	37.10	30.6	75.5	34.7	12.2
Arithmetic	6.06	14.03	23.70	34.23	24.5	57.1	34.7	14.3
Apportionment	5.02	12.71	22.62	34.75	12.2	53.1	32.7	16.3
Ratio I (1870 to 1930)	15.98	31.75	117.74	276.38	44.9	83.7	38.8	28.5
Ratio II (1930)	5.24	12.80	21.14	40.11	20.4	53.1	49.0	28.6
Ratio III (1900 to 1930 modified)	6.11	13.94	23.57	36.61	16.3	55.1	34.7	10.2
Adjusted to census count								
Cohort-survival (T&W)								
With migration	5.16	10.54	22.75	35.69	12.2	38.8	38.8	49.0
No migration	5.84	15.77	26.65	55.35	20.4	51.0	55.1	59.2
Geometric	7.64	13.43	25.15	31.34	30.6	57.1	34.7	28.6
Arithmetic	6.10	10.96	23.81	33.98	22.4	44.9	34.7	38.8
Apportionment	5.06	10.57	22.70	33.61	12.2	40.8	32.7	40.8
Ratio I (1870 to 1930)	15.99	30.87	117.42	310.83	44.9	67.3	38.8	34.7
Ratio II (1930)	5.24	12.59	21.26	34.63	18.4	44.9	46.9	57.1
Ratio III (1900 to 1930 modified)	6.18	10.28	23.68	32.54	16.3	40.8	34.7	36.7

* With migration.

tion of errors of 10 per cent or more, the Ratio III (unadjusted), the cohort-survival (with migration), and the arithmetic (unadjusted) projections are the best.

Apparently the cohort-survival (with migration), the apportionment, and the Ratio III (unadjusted) results make the consistently best show-

ings. The differences between most of the various summary measures, however, are too small to justify any definite conclusions except with regard to the unfavorable showing of Ratio I.

It does appear that projections prepared by any of the methods mentioned above as being consistently best will be better guides for ten and twenty years in the future than the most recently available census data or current estimates. Assuming that the state populations would be the same in 1950 as in 1930 yields an APE of 19; also, 74 per cent of the "projections" are in error by 10 per cent or more. These values are notably higher than those for the cohort-survival (with migration), the apportionment, and the Ratio III (unadjusted) methods.

Control totals.—It has generally been assumed that the best current estimates[6] of the populations of the states are obtained by adjusting the various estimates for the states to add to comparable estimates for the United States. If this is true of current estimates, it would not seem unreasonable to expect it to be true of projections. Hagood and Siegel made this assumption in preparing their article, "Projections of the Regional Distribution of the Population of the United States to 1975;"[7] the method described by them involves adjusting the appropriate ratios to sum exactly to 1.00.

The hypothesis that the use of independent control totals increases the accuracy of state projections can be tested by comparing the summary measures for the unadjusted projections with those for the adjusted projections, both those adjusted to the Thompson-Whelpton national totals and those adjusted to the census counts, for each method.[8] (The constant, apportionment, Ratio I, and Ratio II methods cannot be included in this comparison, of course.) Although the results are somewhat inconclusive, the value of the use of control totals appears to be questionable. For 1950, the APE for each of the methods for which a comparison can be made, with the exception of the cohort-survival method (with migration), is somewhat lower for the unadjusted projections than for the adjusted projections. Even the adjusted Ratio III projections, which involve the use of divisional control totals, have a slightly higher APE than the unadjusted Ratio III projections.

[6] "Estimate" is used here as meaning a figure for a current date which usually does not depend on extrapolation to any major extent.

[7] Margaret Jarman Hagood and Jacob S. Siegel, "Projections of the Regional Distribution of the Population of the United States to 1975," *Agricultural Economics Research*, III (1951), pp. 41–52.

[8] As the base populations are free of error in comparison with the implicit projections of population change, it can be argued that the adjustment of the projections should have been in proportion to the projected change in the population, or some other factor, rather than the projected populations. It may be desirable to include this alternative in future tests of the sort described here.

It is obvious, of course, that adjusting to the actual census count reduces the sum of the differences between the projections and the actual population to zero and, further, that if all states had the same percentage error, proportionate adjustment to the census count would eliminate all errors. However—and this is apparently an important however—the value of a proportionate adjustment depends on the accuracy of the control total and on the distribution of the errors. If the errors are randomly distributed with regard to direction (that is, if the number of positive errors is approximately 50 per cent of all errors), then adjustments can only introduce a bias toward over-estimating or under-estimating. If all of the gross errors are positive or all are negative, and the errors in the projections for the remaining states are negligible, then proportionate adjustment will tend to decrease but not eliminate, the gross errors at the cost of increasing the errors in the projections for the remaining states. Thus, for 1950, adjusting introduced (or added to) a bias towards under-estimating in the Ratio III, geometric, and arithmetic methods. On the other hand, adjusting generally but not consistently yielded better results for 1950 in terms of the maximum per cent error and proportion of errors of ten per cent or more.

The projections adjusted to the actual census count are generally better than those adjusted to a national projection, as might be expected.

Length of projection period.—Inspection of Table 1 shows that the projections for 1940, which involve a 10-year projection period, are subject to smaller errors than the projections for 1950, which involve a 20-year projection period. For all of the methods combined, the APE for 1940 is 7, and 22 per cent of projections are in error by 10 per cent or more; for 1950, these measures are 15 and 54 per cent, respectively. This should be a warning against the claim sometimes made that population projections are satisfactory guides for the long-run trend even when they deviate in the short-run.

Size of population and migration rate.—It seems worth while to investigate the relation of the accuracy of the projections to the size of the base population and to the size of the migration rate characteristic of the area prior to the projection period. Correlation coefficients and regression equations would yield the most useful measures of the relation between these items. However, the measures shown in Table 2 and in Table 3 are indicative of the relations. Table 2 shows summary measures of the errors in the 1950 projections for the 24 states with populations of 1.86 million or more in 1930 and the 25 states with populations of 1.85 million or less in 1930; Table 3 shows

similar measures for the 24 states with 1920–30 average migration rates of 0.5 or less per year (regardless of the direction of the migration) and the 25 states with 1920–30 average migration rates of 0.6 or more per year.[9]

TABLE 2.—SUMMARY OF PER CENT ERRORS OF PROJECTIONS TO 1950 OF THE POPULATIONS OF THE STATES, FOR SELECTED METHODS, BY SIZE OF POPULATION IN 1930

Method	Average per cent error		Maximum per cent error		Proportion of errors of 10 per cent or more (expressed as a per cent)		Proportion of positive errors (expressed as a per cent)	
	24 largest states	25 smallest states	24 largest states	25 smallest states	24 largest states	25 smallest states	24 largest states	25 smallest states
Unadjusted to national total								
Cohort-survival (T&W)								
With migration	8.43	16.45	27.54	39.52	33.3	68.0	12.5	16.0
No migration	10.88	19.17	45.32	43.45	45.8	64.0	37.5	32.0
Geometric	14.44	12.23	47.40	24.70	50.0	56.0	70.8	36.0
Arithmetic	8.74	12.99	38.38	25.86	33.3	68.0	66.7	32.0
Constant	16.12	22.29	46.02	46.89	70.8	76.0	4.2	12.0
Ratio III (1900 to 1930 modified)								
T&W national projection*	8.62	16.15	24.50	34.67	29.2	68.0	16.7	16.0
Census count	7.18	12.38	34.45	28.68	16.7	56.0	62.5	28.0
*Adjusted to T&W national projection**								
Geometric	14.26	19.96	33.51	37.10	75.0	76.0	20.8	4.0
Arithmetic	9.73	18.16	28.13	34.23	37.5	76.0	16.7	12.0
Apportionment	8.37	16.88	26.39	34.75	33.3	72.0	16.7	16.0
Ratio I (1870 to 1930)	19.36	43.64	105.50	276.38	79.2	88.0	12.5	40.0
Ratio II (1930)	8.73	16.71	39.13	40.11	37.5	68.0	33.3	24.0
Ratio III (1900 to 1930 modified)	9.79	17.93	26.41	36.61	37.5	72.0	8.3	12.0
Adjusted to census count								
Cohort-survival (T&W)								
With migration	7.55	13.41	35.69	33.98	20.8	56.0	66.7	32.0
No migration	12.33	19.07	51.80	55.35	37.5	64.0	66.7	52.0
Geometric	11.62	15.17	31.23	31.34	45.8	68.0	33.3	24.0
Arithmetic	7.71	14.08	33.98	28.21	16.7	72.0	54.2	24.0
Apportionment	7.26	13.75	33.61	28.36	20.8	60.0	58.3	24.0
Ratio I (1870 to 1930)	15.12	45.99	124.31	310.83	50.0	84.0	20.8	48.0
Ratio II (1930)	9.45	15.62	33.56	34.63	33.3	56.0	70.8	44.0
Ratio III (1900 to 1930 modified)	6.72	13.71	32.54	30.81	16.7	64.0	50.0	24.0

* With migration.

[9] For this purpose, the absolute values of the rates were used. They were obtained from Henry S. Shryock, Jr., "Internal Migration and the War" (*Journal of the American Statistical Association*, 38 (1943), pp. 16–30).

Table 2 suggests definitely that, for a given method and a given length of projection period, the errors of projections tend to be larger

TABLE 3.—SUMMARY OF PER CENT ERRORS OF PROJECTIONS TO 1950 OF THE POPULATIONS OF THE STATES, FOR SELECTED METHODS, BY AVERAGE MIGRATION RATE FOR 1920–1930

Method	Average per cent error		Maximum per cent error		Proportion of errors of 10 per cent or more (expressed as a per cent)		Proportion of positive errors (expressed as a per cent)	
	[illegible] states with smallest rates	25 states with largest rates	24 states with smallest rates	25 states with largest rates	24 states with smallest rates	25 states with largest rates	24 states with smallest rates	25 states with largest rates
Unadjusted to national total								
Cohort-survival (T&W)								
With migration	11.07	13.92	39.52	32.87	41.7	60.0	12.5	16.0
No migration	12.79	17.34	40.17	45.32	50.0	60.0	20.8	48.0
Geometric	11.61	14.95	47.40	44.56	45.8	60.0	62.5	44.0
Arithmetic	9.42	12.33	38.38	25.01	37.5	64.0	62.5	36.0
Constant	19.24	19.30	42.64	46.89	79.2	68.0	4.2	12.0
Ratio III (1900 to 1930 modified)								
T&W national projection*	10.99	13.88	34.67	31.89	41.7	56.0	16.7	16.0
Census count	8.43	11.18	34.45	25.66	20.8	52.0	54.2	36.0
*Adjusted to T&W national projection**								
Geometric	15.15	19.10	35.76	37.10	75.0	76.0	8.3	16.0
Arithmetic	11.76	16.21	34.23	33.48	50.0	64.0	12.5	16.0
Apportionment	11.00	14.36	34.75	31.57	50.0	56.0	8.3	24.0
Ratio I (1870 to 1930)	24.91	38.32	105.50	276.38	79.2	88.0	20.8	32.0
Ratio II (1930)	10.85	14.68	35.32	40.11	50.0	56.0	20.8	36.0
Ratio III (1900 to 1930 modified)	12.56	15.26	36.61	33.61	50.0	60.0	4.2	16.0
Adjusted to census count								
Cohort-survival (T&W)								
With migration	8.93	12.09	35.69	26.72	29.2	48.0	54.2	44.0
No migration	11.54	19.83	51.80	55.35	33.3	68.0	54.2	64.0
Geometric	10.76	16.00	31.23	31.34	41.7	72.0	29.2	28.0
Arithmetic	8.90	12.94	33.98	27.40	25.0	64.0	45.8	32.0
Apportionment	8.68	12.39	33.61	26.86	20.8	60.0	50.0	32.0
Ratio I (1870 to 1930)	22.58	38.82	124.31	310.83	58.3	76.0	33.3	36.0
Ratio II (1930)	9.[illegible]6	15.41	31.07	34.63	37.5	52.0	54.2	60.0
Ratio III (1900 to 1930 modified)	8.63	11.88	32.54	27.54	25.0	56.0	41.7	32.0

* With migration.

as the size of the population on which the projections are based becomes smaller. Thus, for all methods shown in Table 2, the APE of the

projections for 1950 for the 24 states with the larger populations is 11, while the APE for the other 25 states is 19. Also, 39 per cent of the projections of the larger populations are in error by 10 per cent or more, while 68 per cent of the projections of the smaller populations are in error by 10 per cent or more.

The errors also tend to be larger for the states with the larger average migration rates, according to Table 3. The APE for the 24 states with the smaller migration rates is 12, while the APE for the other states is 17. Errors of 10 per cent or more occur in 45 per cent of the projections for the first 24 states and in 63 per cent of the projections for the second 25 states.

As might be expected, the projections for the 14 states which are both among the 24 states with larger populations and among the 24 states with the smaller migration rates, have a smaller average per cent error (10) and a smaller proportion (33 per cent) of errors of 10 per cent or more than either of the two complete groups.

Need for additional research.—Even though the results of this test are inconclusive, they are probably of sufficient value and interest to warrant additional research along the same general lines. Other areas needing additional research are numerous. A few of these areas are mentioned below. Failure to include the logistic method is one of the more obvious gaps of this study. The question of using controls at any level and the question of a single national control versus controls at the national and various intermediate levels, should be explored further. In the light of the results for the several ratio methods, the best balance in the basic assumptions between emphasis on long-run trend and emphasis on recent experience, should be investigated. In connection with this, some measures relating to the possibility of allowing specifically for various economic conditions in the future by the use of "representative" trends from appropriate past periods should be obtained. Because any period represents unique experience, projections for additional combinations of periods should be studied. In addition, the significance of the various measures should be tested.

Conclusions.—Although not one of the methods tested is clearly superior to the others, the cohort-survival (with migration), the apportionment, and the Ratio III (unadjusted) results make the consistently best showings on the basis of average per cent error and proportion of errors of 10 per cent or more.

The following hypotheses, while not proven by this test, are consistent with the results obtained:

1. Projections obtained by these three methods will be better guides

for ten and twenty years in the future than the most recent data on current population size.

2. The value of the use of independent control totals is questionable.

3. The errors of projections tend to increase almost directly as the length of the projection period increases.

4. The errors of projections tend to be larger for areas with smaller base populations.

5. The errors of projections tend to be larger for areas with larger net migration rates in the recent past.[10]

APPENDIX

Discussions of the accuracy of population projections will be found in the following items:

Davis, J. S., *The Population Upsurge in the United States*, War-Peace Pamphlets No. 12, Stanford: Food Research Institute, Stanford University, 1949.

Dorn, Harold F., "Pitfalls in Population Forecasts and Projections," *Journal of the American Statistical Association*, 45 (1950), 311–34.

Hotelling, Harold, "Differential Equations Subject to Error and Population Estimates," *Journal of the American Statistical Association*, 22 (1927), 283–314.

Myers, Robert J., "Comparison of Demographic Rates Assumed by the National Resources Committee with Actual Experience," *Journal of the American Statistical Association*, 38 (1943), 201–09.

Schmitt, Robert C., and Albert H. Crosetti, "Accuracy of the Ratio Method for Forecasting City Population," *Land Economics*, XXVII (1951), 346–48.

Schultz, Henry, "The Standard Error of a Forecast from a Curve," *Journal of the American Statistical Association*, 25 (1930), 139–85.

Shryock, Henry S., Jr., "Forecasts of Population in the United States," *Population Studies*, III (1950), 406–12.

Wilson, Edwin B., and Ruth R. Puffer, "Least Squares and Laws of Population Growth," *Proceedings of the American Academy of Arts and Sciences*, 68 (1933), 285–382.

Descriptions of various methods of projecting population, references to pertinent literature, comments on the development of projections, and a priori evaluations of various methods, will be found in the following items:

Hagood, Margaret Jarman, and Jacob S. Siegel, "Projections of the Regional

---

[10] An article by Robert C. Schmitt and Albert H. Crosetti entitled "Accuracy of the Ratio Method for Forecasting City Population" (*Land Economics*, XXVII (1951), pp. 346–48), has just come to the attention of the author. This article describes a test of the accuracy of the ratio method in predicting the population of selected large cities and of variations in accuracy with length of projection period, size of population, and growth rate. The findings of Schmitt and Crosetti are in agreement with (4) above but not with (5) above. It is possible that the results would agree if coefficients of partial correlation had been used to measure the association of accuracy or projections, size of base population, and growth rates or migration rates.

Distribution of the Population of the United States to 1975," *Agricultural Economics Research*, III (1951), 41–52.

Notestein, Frank W., et al., *The Future Population of Europe and the Soviet Union*, Geneva: League of Nations, 1944, 199–234.

Reed, Lowell J., "Population Growth and Forecasts," *The Annals of the American Academy of Political and Social Science*, 188 (1936), 159–66.

Taeuber, Irene B., "The Development of Population Predictions in Europe and the Americas," *Estadistica*, 7 (1944), 323–46.

Taeuber, Irene B., "Current Items—Literature on Future Populations, 1943–1948," *Population Index*, 15 (1949), 2–30.

Whelpton, P. K., *Needed Population Research*, Lancaster, Pennsylvania: The Science Press Printing Co., 1938, 1–11.

APPENDIX TABLE A.—PROJECTIONS TO 1940 OF THE 1930 POPULATIONS OF THE STATES, BY SELECTED METHODS

(In thousands. Each figure has been independently rounded)

State	1940 enumerated population	Unadjusted to national total						
		Cohort-survival (T&W)		Geometric	Arithmetic	Constant	Ratio III (1900 to 1930 modified)	
		With migration	No migration				T&W national projection*	Census count
United States	131,669	131,865	132,098	143,401	139,423	122,775	133,011	132,818
Alabama	2,833	2,801	3,024	2,978	2,937	2,646	2,767	2,763
Arizona	499	513	491	564	535	436	527	527
Arkansas	1,949	1,898	2,113	1,960	1,954	1,854	1,881	1,878
California	6,907	6,808	5,810	9,290	7,873	5,677	7,915	7,903
Colorado	1,123	1,082	1,104	1,139	1,130	1,036	1,084	1,082
Connecticut	1,709	1,726	1,682	1,863	1,828	1,607	1,753	1,751
Delaware	267	248	249	254	253	238	242	241
District of Col.	663	513	488	540	535	487	512	512
Florida	1,897	1,716	1,557	2,203	1,956	1,468	1,859	1,856
Georgia	3,124	2,915	3,308	2,921	2,921	2,909	2,877	2,873
Idaho	525	455	504	458	458	445	453	453
Illinois	7,897	8,177	7,933	8,943	8,748	7,631	8,291	8,278
Indiana	3,428	3,405	3,398	3,570	3,539	3,239	3,328	3,323
Iowa	2,538	2,497	2,649	2,538	2,536	2,471	2,441	2,438
Kansas	1,801	1,940	2,035	1,997	1,990	1,881	1,887	1,884
Kentucky	2,846	2,732	2,955	2,823	2,808	2,615	2,665	2,661
Louisiana	2,364	2,265	2,357	2,446	2,397	2,102	2,261	2,258
Maine	847	823	852	827	826	797	786	785
Maryland	1,821	1,730	1,710	1,831	1,809	1,632	1,724	1,721
Massachusetts	4,317	4,458	4,398	4,677	4,637	4,250	4,423	4,416
Michigan	5,256	5,550	5,230	6,349	5,988	4,842	5,653	5,645
Minnesota	2,792	2,647	2,767	2,749	2,736	2,564	2,598	2,595
Mississippi	2,184	2,137	2,293	2,033	2,224	2,010	2,085	2,082
Missouri	3,785	3,709	3,815	3,864	3,849	3,629	3,645	3,639
Montana	559	528	588	527	527	538	535	534
Nebraska	1,316	1,415	1,501	1,463	1,458	1,378	1,384	1,382
Nevada	110	92	95	107	104	91	98	98
New Hampshire	492	476	480	488	487	465	463	463
New Jersey	4,160	4,488	4,218	5,144	4,905	4,041	4,697	4,690
New Mexico	532	466	496	495	485	423	457	456
New York	13,479	13,709	12,989	15,187	14,737	12,588	13,798	13,778
North Carolina	3,572	3,575	3,682	3,907	3,767	3,170	3,521	3,515
North Dakota	642	702	790	716	714	681	680	679
Ohio	6,908	7,107	6,956	7,044	7,512	6,647	7,163	7,153
Oklahoma	2,336	2,620	2,785	2,819	2,755	2,396	2,611	2,607
Oregon	1,090	1,028	978	1,156	1,120	954	1,081	1,079
Pennsylvania	9,900	10,086	10,203	10,612	10,520	9,631	10,099	10,084
Rhode Island	713	732	717	780	769	687	725	724
South Carolina	1,900	1,781	2,000	1,794	1,792	1,739	1,758	1,755
South Dakota	643	728	783	753	748	693	706	705
Tennessee	2,916	2,754	2,937	2,920	2,888	2,617	2,719	2,715
Texas	6,415	6,469	6,506	7,236	6,958	5,825	6,587	6,577
Utah	550	559	584	572	565	508	539	538
Vermont	359	365	379	367	367	360	340	340
Virginia	2,678	2,495	2,682	2,537	2,532	2,422	2,454	2,450
Washington	1,736	1,658	1,620	1,795	1,765	1,563	1,693	1,691
West Virginia	1,902	1,913	1,979	2,035	1,988	1,729	1,908	1,905
Wisconsin	3,138	3,123	3,195	3,273	3,238	2,939	3,093	3,088
Wyoming	251	247	251	261	256	226	245	244

* With migration.

APPENDIX TABLE A.—*Continued*

(In thousands. Each figure has been independently rounded)

State	Adjusted to T&W national projection*					
	Geometric	Arithmetic	Apportion-ment	Ratio I (1870 to 1930)	Ratio II (1930)	Ratio III (1900 to 1930 modified)
United States	131,865	131,865	131,865	131,865	131,865	131,865
Albama	2,734	2,778	2,805	2,664	2,842	2,751
Arizona	519	506	490	673	468	520
Arkansas	1,802	1,848	1,909	1,965	1,992	1,869
California	8,543	7,446	6,875	6,989	6,098	7,699
Colorado	1,047	1,068	1,087	1,477	1,112	1,068
Connecticut	1,713	1,729	1,727	1,648	1,726	1,749
Delaware	234	240	247	224	256	239
District of Col.	497	506	513	514	523	507
Florida	2,026	1,850	1,734	1,741	1,577	1,839
Georgia	2,686	2,763	2,915	2,888	3,124	2,845
Idaho	421	433	452	646	478	447
Illinois	8,223	8,274	8,240	7,820	8,196	8,256
Indiana	3,283	3,347	3,403	3,099	3,478	3,314
Iowa	2,334	2,399	2,507	2,400	2,654	2,428
Kansas	1,836	1,882	1,940	2,083	2,020	1,877
Kentucky	2,596	2,656	2,720	2,519	2,808	2,650
Louisiana	2,250	2,267	2,263	2,136	2,257	2,247
Maine	761	781	813	712	856	784
Maryland	1,684	1,711	1,728	1,582	1,752	1,705
Massachusetts	4,300	4,386	4,461	4,338	4,564	4,411
Michigan	5,838	5,663	5,467	5,195	5,201	5,629
Minnesota	2,528	2,588	2,658	2,901	2,754	2,585
Mississippi	1,870	2,103	2,127	1,991	2,159	2,073
Missouri	3,553	3,641	3,749	3,521	3,898	3,625
Montana	484	498	538	765	577	528
Nebraska	1,345	1,379	1,421	1,727	1,480	1,376
Nevada	98	99	98	92	98	97
New Hampshire	449	461	477	422	500	462
New Jersey	4,730	4,639	4,513	4,391	4,341	4,654
New Mexico	455	458	457	462	455	450
New York	13,965	13,938	13,761	12,804	13,520	13,674
North Carolina	3,593	3,562	3,496	3,244	3,405	3,482
North Dakota	658	675	699	1,398	731	676
Ohio	7,029	7,105	7,119	6,620	7,139	7,133
Oklahoma	2,592	2,605	2,592	3,521	2,573	2,594
Oregon	1,063	1,059	1,044	1,174	1,024	1,051
Pennsylvania	9,758	9,950	10,116	9,732	10,344	10,008
Rhode Island	717	727	732	712	738	723
South Carolina	1,650	1,695	1,768	1,727	1,867	1,738
South Dakota	692	707	723	1,108	744	702
Tennessee	2,685	2,732	2,765	2,532	2,810	2,703
Texas	6,654	6,581	6,443	6,791	6,256	6,545
Utah	526	534	539	580	545	531
Vermont	337	347	363	316	386	339
Virginia	2,333	2,395	2,482	2,334	2,601	2,427
Washington	1,651	1,669	1,673	2,545	1,679	1,647
West Virginia	1,871	1,880	1,871	1,846	1,857	1,887
Wisconsin	3,010	3,063	3,102	2,980	3,157	3,080
Wyoming	240	242	242	316	242	241

* With migration.

APPENDIX TABLE A.—*Continued*

(In thousands. Each figure has been independently rounded)

State	Adjusted to census count								
	Cohort-survival (T&W)		Geometric	Arithmetic	Apportionment	Ratio I (1870 to 1930)	Ratio II (1930)	Ratio III (1900 to 1930 modified)
	With migration	No migration						
United States	131,669	131,669	131,669	131,669	131,669	131,669	131,669	131,669
Alabama	2,797	3,014	2,730	2,774	2,802	2,660	2,838	2,747
Arizona	512	489	518	505	488	672	467	519
Arkansas	1,895	2,106	1,800	1,846	1,908	1,962	1,989	1,866
California	6,798	5,790	8,530	7,435	6,849	6,978	6,089	7,687
Colorado	1,080	1,100	1,046	1,067	1,086	1,475	1,111	1,067
Connecticut	1,723	1,676	1,711	1,726	1,725	1,646	1,723	1,746
Delaware	248	248	234	239	246	224	256	239
District of Col.	512	486	496	505	513	514	522	506
Florida	1,714	1,552	2,023	1,847	1,729	1,738	1,575	1,836
Georgia	2,911	3,297	2,682	2,758	2,915	2,884	3,119	2,841
Idaho	454	502	421	432	452	645	477	446
Illinois	8,165	7,906	8,211	8,262	8,227	7,808	8,183	8,243
Indiana	3,400	3,387	3,278	3,342	3,399	3,094	3,473	3,309
Iowa	2,493	2,640	2,330	2,395	2,506	2,396	2,650	2,425
Kansas	1,937	2,028	1,833	1,879	1,939	2,080	2,017	1,874
Kentucky	2,728	2,945	2,592	2,652	2,718	2,515	2,804	2,646
Louisiana	2,262	2,349	2,246	2,264	2,259	2,133	2,254	2,244
Maine	822	849	760	780	813	711	855	783
Maryland	1,727	1,704	1,681	1,708	1,726	1,580	1,750	1,702
Massachusetts	4,452	4,383	4,294	4,379	4,457	4,332	4,557	4,404
Michigan	5,542	5,212	5,829	5,655	5,454	5,188	5,193	5,621
Minnesota	2,643	2,758	2,524	2,584	2,656	2,897	2,750	2,581
Mississippi	2,134	2,285	1,867	2,100	2,124	1,988	2,155	2,070
Missouri	3,704	3,802	3,547	3,635	3,747	3,516	3,892	3,620
Montana	527	586	484	497	538	764	577	527
Nebraska	1,413	1,496	1,343	1,377	1,420	1,725	1,478	1,374
Nevada	92	95	98	99	98	92	98	97
New Hampshire	475	478	448	460	477	421	499	462
New Jersey	4,481	4,204	4,723	4,632	4,503	4,385	4,334	4,648
New Mexico	465	494	455	458	456	461	454	450
New York	13,689	12,945	13,944	13,918	13,735	12,785	13,500	13,653
North Carolina	3,750	3,670	3,587	3,557	3,489	3,239	3,400	3,477
North Dakota	701	787	657	674	699	1,396	730	675
Ohio	7,097	6,932	7,019	7,095	7,109	6,610	7,128	7,123
Oklahoma	2,616	2,776	2,588	2,602	2,588	3,516	2,570	2,590
Oregon	1,027	975	1,061	1,058	1,043	1,172	1,023	1,060
Pennsylvania	10,071	10,169	9,744	9,935	10,106	9,717	10,329	9,993
Rhode Island	731	715	716	726	731	711	737	722
South Carolina	1,778	1,993	1,647	1,693	1,767	1,725	1,865	1,736
South Dakota	727	780	691	706	722	1,106	743	[illegible]
Tennessee	2,750	2,972	2,681	2,728	2,762	2,528	2,806	2,699
Texas	6,460	6,484	6,644	6,571	6,430	6,781	6,247	6,535
Utah	558	582	525	533	538	579	545	530
Vermont	364	378	337	346	363	316	386	339
Virginia	2,491	2,673	2,330	2,391	2,481	2,331	2,597	2,423
Washington	1,656	1,615	1,649	1,667	1,671	2,541	1,677	1,644
West Virginia	1,910	1,972	1,868	1,878	1,867	1,843	1,854	1,884
Wisconsin	3,118	3,184	3,005	3,058	3,099	2,976	3,152	3,075
Wyoming	247	250	239	242	242	316	242	241

APPENDIX TABLE A.—*Continued*

(In thousands. Each figure has been independently rounded)

State	Estimated 1950 de jure population	Unadjusted to national total						
		Cohort-survival (T&W)		Geometric	Arithmetic	Constant	Ratio III (1900 to 1930 modified)	
		With migration	No migration				T&W national projection*	Census count
United States	151,116	138,442	139,542	169,729	153,071	122,775	141,209	154,136
Alabama	3,088	2,921	3,413	3,341	3,228	2,646	2,845	3,105
Arizona	750	580	541	731	633	436	608	664
Arkansas	1,932	1,925	2,366	2,072	2,054	1,854	1,889	2,062
California	10,517	7,621	5,751	15,203	10,068	5,677	10,188	11,121
Colorado	1,323	1,111	1,160	1,253	1,223	1,036	1,115	1,217
Connecticut	2,019	1,814	1,724	2,161	2,048	1,607	1,864	2,035
Delaware	320	254	255	272	268	238	243	265
District of Col.	778	522	470	600	583	467	530	578
Florida	2,764	1,904	1,620	3,307	2,443	1,468	2,220	2,423
Georgia	3,443	2,913	3,709	2,933	2,933	2,909	2,833	3,092
Idaho	594	465	562	472	471	445	457	498
Illinois	8,738	8,509	8,015	10,480	9,866	7,631	8,783	9,587
Indiana	3,966	3,522	3,509	3,936	3,840	3,239	3,375	3,684
Iowa	2,644	2,502	2,799	2,607	2,602	2,471	2,402	2,622
Kansas	1,908	1,973	2,159	2,120	2,099	1,881	1,878	2,050
Kentucky	2,944	2,851	3,314	3,049	3,001	2,615	2,687	2,933
Louisiana	2,699	2,396	2,599	2,848	2,693	2,102	2,377	2,595
Maine	924	841	903	858	855	797	772	843
Maryland	2,327	1,300	1,759	2,055	1,986	1,632	1,786	1,950
Massachusetts	4,718	4,597	4,454	5,147	5,025	4,250	4,531	4,945
Michigan	6,412	6,140	5,511	8,324	7,133	4,842	6,335	6,914
Minnesota	3,008	2,693	2,931	2,948	2,909	2,564	2,607	2,846
Mississippi	2,187	2,239	2,570	2,057	2,438	2,010	2,131	2,326
Missouri	3,989	3,730	3,930	4,113	4,069	3,629	3,631	3,964
Montana	595	517	627	560	516	538	530	578
Nebraska	1,335	1,434	1,602	1,552	1,537	1,378	1,378	1,505
Nevada	159	96	97	125	118	91	104	113
New Hampshire	538	485	491	512	509	465	459	501
New Jersey	4,842	4,813	4,285	6,548	5,769	4,041	5,244	5,724
New Mexico	675	510	568	580	546	423	482	526
New York	14,909	14,398	12,993	18,322	16,886	12,588	14,718	16,066
North Carolina	4,056	3,952	4,246	4,815	4,363	3,170	3,794	4,141
North Dakota	625	716	897	752	747	681	674	736
Ohio	8,003	7,433	7,127	8,791	8,378	6,647	7,541	8,231
Oklahoma	2,250	2,797	3,154	3,317	3,114	2,396	2,772	3,025
Oregon	1,533	1,073	980	1,400	1,286	954	1,183	1,292
Pennsylvania	10,598	10,376	10,650	11,693	11,410	9,631	10,406	11,358
Rhode Island	782	765	733	884	850	687	751	820
South Carolina	2,117	1,822	2,291	1,851	1,846	1,739	1,760	1,921
South Dakota	656	757	868	817	803	693	712	777
Tennessee	3,314	2,853	3,244	3,260	3,160	2,617	2,782	3,037
Texas	7,674	6,974	7,101	8,990	8,091	5,825	7,199	7,858
Utah	694	610	663	645	622	508	560	612
Vermont	383	368	397	374	374	360	323	353
Virginia	3,260	2,552	2,952	2,658	2,642	2,422	2,462	2,687
Washington	2,341	1,712	1,632	2,062	1,967	1,563	1,790	1,954
West Virginia	2,032	2,088	2,238	2,394	2,247	1,729	2,045	2,232
Wisconsin	3,464	3,253	3,408	3,645	3,538	2,939	3,196	3,488
Wyoming	285	263	271	301	286	226	259	282

* With migration.

State	Adjusted to T&W national projection*					
	Geometric	Arithmetic	Apportionment	Ratio I (1870 to 1930)	Ratio II (1930)	Ratio III (1900 to 1930 modified)
United States	138,442	138,442	138,442	138,442	138,442	138,442
Alabama	2,725	2,863	2,920	2,617	2,984	2,809
Arizona	596	562	529	914	491	590
Arkansas	1,690	1,822	1,948	2,021	2,091	1,862
California	12,400	8,931	7,742	8,030	6,402	9,604
Colorado	1,022	1,085	1,124	1,883	1,168	1,082
Connecticut	1,762	1,817	1,815	1,647	1,812	1,850
Delaware	221	238	252	222	269	237
District of Col.	489	517	532	526	549	516
Florida	2,697	2,167	1,927	1,938	1,656	2,164
Georgia	2,393	2,602	2,920	2,824	3,280	2,761
Idaho	385	418	457	831	502	448
Illinois	8,549	8,751	8,682	7,822	8,604	8,689
Indiana	3,211	3,406	3,521	2,949	3,652	3,339
Iowa	2,126	2,308	2,532	2,298	2,786	2,374
Kansas	1,729	1,862	1,984	2,215	2,121	1,856
Kentucky	2,487	2,662	2,796	2,395	2,948	2,653
Louisiana	2,323	2,389	2,380	2,132	2,370	2,344
Maine	700	758	824	609	899	766
Maryland	1,676	1,762	1,798	1,523	1,840	1,741
Massachusetts	4,198	4,457	4,614	4,319	4,792	4,496
Michigan	6,789	6,327	5,919	5,385	5,460	6,267
Minnesota	2,404	2,580	2,726	3,129	2,891	2,577
Mississippi	1,678	2,162	2,211	1,952	2,266	2,104
Missouri	3,355	3,609	3,836	3,392	4,092	3,589
Montana	457	457	538	983	606	514
Nebraska	1,266	1,364	1,453	1,994	1,554	1,362
Nevada	102	104	104	97	103	101
New Hampshire	418	451	486	388	525	455
New Jersey	5,341	5,117	4,854	4,596	4,557	5,145
New Mexico	478	484	481	485	477	467
New York	14,944	14,979	14,609	12,751	14,194	14,440
North Carolina	3,927	3,870	3,731	3,240	3,575	3,698
North Dakota	614	663	712	2,354	768	666
Ohio	7,170	7,432	7,461	6,479	7,495	7,460
Oklahoma	2,705	2,762	2,738	4,624	2,702	2,732
Oregon	1,142	1,141	1,110	1,357	1,075	1,116
Pennsylvania	9,537	10,121	10,467	9,636	10,860	10,209
Rhode Island	721	754	764	720	775	745
South Carolina	1,510	1,688	1,789	1,689	1,961	1,716
South Dakota	667	712	745	1,561	781	703
Tennessee	2,659	2,803	2,872	2,437	2,950	2,747
Texas	7,332	7,177	6,890	7,490	6,568	7,096
Utah	526	552	561	623	573	544
Vermont	305	331	366	277	406	32[illegible]
Virginia	2,168	2,343	2,525	2,229	2,731	2,399
Washington	1,682	1,745	1,753	3,627	1,763	1,687
West Virginia	1,953	1,993	1,973	1,897	1,950	1,993
Wisconsin	2,973	3,188	3,221	2,949	3,314	3,161
Wyoming	246	254	254	401	254	251

* With migration.

State	Adjusted to census count							
	Cohort-survival (T&W)		Geometric	Arithmetic	Apportionment	Ratio I (1870 to 1930)	Ratio II (1930)	Ratio III (1900 to 1930 modified)
	With migration	No migration						
United States	151,116	151,116	151,116	151,116	151,116	151,116	151,116	151,116
Alabama	3,188	3,696	2,975	3,125	3,141	2,856	3,257	3,066
Arizona	633	586	650	613	604	997	536	644
Arkansas	2,101	2,562	1,844	1,989	2,024	2,206	2,283	2,032
California	8,319	6,228	13,536	9,749	9,412	8,765	6,988	10,483
Colorado	1,213	1,256	1,115	1,185	1,195	2,055	1,275	1,181
Connecticut	1,980	1,867	1,924	1,983	1,982	1,798	1,978	2,019
Delaware	277	276	242	260	264	242	293	258
District of Col.	570	509	534	565	569	574	599	563
Florida	2,078	1,754	2,944	2,366	2,298	2,116	1,807	2,362
Georgia	3,180	4,017	2,612	2,840	2,930	3,083	3,580	3,014
Idaho	508	609	420	456	467	907	548	483
Illinois	9,288	8,680	9,331	9,552	9,532	8,538	9,392	9,484
Indiana	3,844	3,800	3,504	3,718	3,750	3,219	3,986	3,645
Iowa	2,731	3,031	2,321	2,519	2,582	2,509	3,041	2,591
Kansas	2,154	2,338	1,887	2,032	2,066	2,418	2,315	2,026
Kentucky	3,112	3,589	2,714	2,906	2,943	2,614	3,218	2,896
Louisiana	2,615	2,815	2,536	2,607	2,605	2,327	2,587	2,558
Maine	918	978	764	828	846	665	981	836
Maryland	1,965	1,905	1,829	1,923	1,933	1,662	2,008	1,901
Massachusetts	5,018	4,824	4,582	4,865	4,909	4,715	5,231	4,907
Michigan	6,702	5,968	7,411	6,906	6,791	5,878	5,960	6,840
Minnesota	2,940	3,174	2,624	2,817	2,857	3,415	3,156	2,818
Mississippi	2,444	2,783	1,831	2,360	2,374	2,131	2,474	2,297
Missouri	4,072	4,256	3,662	3,940	4,003	3,702	4,467	3,918
Montana	564	679	498	499	538	1,073	662	561
Nebraska	1,565	1,735	1,382	1,488	1,513	2,176	1,696	1,487
Nevada	105	105	111	114	114	106	112	110
New Hampshire	529	532	456	492	502	423	573	497
New Jersey	5,254	4,640	5,830	5,586	5,511	5,017	4,974	5,616
New Mexico	557	615	516	529	528	529	521	510
New York	15,716	14,071	16,313	16,350	16,244	13,918	15,494	15,762
North Carolina	4,314	4,598	4,287	4,224	4,185	3,536	3,902	4,037
North Dakota	782	971	670	723	737	2,569	838	727
Ohio	8,114	7,718	7,827	8,112	8,119	7,072	8,181	8,143
Oklahoma	3,053	3,416	2,953	3,015	3,006	5,047	2,949	2,982
Oregon	1,171	1,061	1,247	1,245	1,237	1,481	1,174	1,218
Pennsylvania	11,326	11,534	10,411	11,047	11,144	10,518	11,855	11,143
Rhode Island	835	794	787	823	825	786	846	813
South Carolina	1,989	2,481	1,638	1,788	1,830	1,844	2,140	1,873
South Dakota	826	940	728	777	786	1,692	853	768
Tennessee	3,114	3,513	2,902	3,060	3,079	2,660	3,221	2,999
Texas	7,613	7,690	8,004	7,834	7,752	8,175	7,169	7,746
Utah	666	717	574	602	605	680	625	593
Vermont	402	430	333	362	372	302	443	350
Virginia	2,786	3,197	2,366	2,558	2,609	2,433	2,981	2,619
Washington	1,869	1,767	1,836	1,904	1,907	3,959	1,924	1,842
West Virginia	2,279	2,424	2,131	2,176	2,170	2,079	2,128	2,176
Wisconsin	3,551	3,691	3,245	3,426	3,448	3,219	3,617	3,451
Wyoming	287	293	268	277	277	438	278	274

FACTORS IN INTERPRETING MORTALITY AFTER RETIREMENT

ROBERT J. MYERS
Social Security Administration

Currently there is considerable discussion as to the effect of compulsory retirement on the national economy and on the vitality and longevity of the individuals concerned. Some experiences would seem to indicate that retirement causes higher mortality than is standard for the ages concerned. Frequently, however, such conclusions are not warranted because the individuals who do retire under voluntary provisions tend to be those who are in poor health. When retirement is compulsory, such experience as is available does not indicate high mortality but this is probably, at least in part, due to the fact that many quite healthy workers are among the retired group in contrast with the situation under plans having voluntary retirement. There is no conclusive data currently on hand to indicate for a given group of individuals what the effect of retirement on mortality really is depending upon whether similar groups of individuals could retire or could continue working.

IN RECENT years considerable discussion has been given to the advantages of continuing individuals in employment beyond age 65 rather than having compulsory retirement at that age, as is the case in many retirement plans. Such advantages are said to accrue both to the individual involved and to the nation.

One of the advantages frequently claimed insofar as the individual is concerned is that a person compelled to retire loses his vitality and thus tends to die much earlier than if allowed to continue in gainful employment. This runs contrary to the viewpoint frequently expressed many years ago that workers were being compelled to remain at work because there was no pension plan to take care of them so that their inevitable end was death from exhaustion. Instead, it was advocated that such workers should be allowed to spend their declining years in peace and leisure while supported by a pension.

Currently, there are about 15,000 pension plans supplementary to the social security program. Many of these, following to some extent previous employer practice, provide for a compulsory retirement age (often at 65). In the majority of plans, retirement may be deferred with the consent of the employer. That retirement at age 65 is by no means universal is indicated by the fact that the average retirement age under the Old-Age and Survivors Insurance program is currently 69

for men and somewhat over 68 for women (in 1940–50 it was generally about one year higher).

This paper will examine the question as to the effect of retirement on mortality. Before proceeding further, let me issue the warning that no clear and definite conclusions will be or can be drawn because there are so many conflicting factors involved.

Unfortunately, specific and reliable data on this subject are not available. The analysis is complicated by the question as to whether people retire because they are disabled and are thus subject to high mortality, or on the other hand whether the high mortality is the result of retirement. Data as to the mortality of retired persons will be examined for several governmental retirement systems and for a few non-governmental pension plans in an effort to throw some light on this matter.

EXPERIENCE TO BE EXPECTED IN VARIOUS TYPES OF PLANS

Before proceeding to such actual data as are available, it will be worthwhile to examine briefly the effect that the particular provisions of the plan might have on the resulting experience. This is an extremely important factor because completely different results may be obtained for what is essentially the same underlying mortality—all depending on the structure of the benefits provided and the administrative procedure adopted.

In considering various possible hypothetical pension plans, let it first be assumed that mortality is not affected by retirement. Then we shall be able to see that any indications of lower or higher mortality following retirement arise solely from the particular plan and its provisions.

First, consider a plan which has no benefits payable before age 65—either early age retirements or disability retirements—but which has compulsory retirement at age 65 and which pays an annuity beginning at age 65 to those who previously left service because of disability. Under this plan, mortality after age 65 would, for the entire retired group, be fairly comparable with that previous to age 65, or with what might be termed the "general level." Of course, as between those who were in active service when they attained age 65 and those disabled persons previously separated from service who receive an annuity at age 65, the former would experience lower mortality.

Second, consider what the situation would be if the previous plan did not have compulsory retirement at age 65. For ages shortly after age 65, it is likely that the mortality experience would be higher than the general level because there would be a tendency for the less healthy lives to retire at or shortly after age 65 and for the healthier lives to

continue at work. After age 70, the mortality experience of the total retired would approach the general level of mortality because virtually everybody would have retired by then.

Third, consider the case where disability pensions are provided (or where disabled persons receive no vested rights for a pension at age 65). If retirement is compulsory at age 65, the experience for non-disabled retired workers will show definitely lower mortality than the general level at the ages shortly after age 65 but eventually will merge into the general level. If retirement is not compulsory at age 65, the resulting mortality experience will probably be somewhat higher than the general level at the ages just beyond age 65 and not as high as for the group of disabled pensioners.

Fourth, consider the experience under a plan which permits optional retirement before age 65. There is a subdivision between disability pensioners and others (as there well might be because of a differential in benefit amount favoring the former). The disability pensioners will experience quite high mortality, while the other pensioners, at least for a few years, will experience very low mortality. This latter group would undoubtedly obtain the larger disability pensions if possible and therefore must be considered to be quite select medically.

EXPERIENCE UNDER OLD-AGE AND SURVIVORS INSURANCE PROGRAM

The old-age and survivors insurance program covers some 80% of the paid civilian jobs in the country. In its actual operation, a vast amount of valuable mortality experience has been accumulated. Unfortunately, it has not been possible to tabulate and analyze all of this vast store of information, especially in regard to mortality data stratified by duration of retirement.

In the early 1940's a brief investigation was made as to select mortality by age and duration of retirement. This indicated that, as contrasted with general population mortality, a person who has just retired has about 15% higher mortality. But this differential rapidly diminishes until after two or three years it has virtually disappeared.

More recently it has been possible to make an investigation as to the over-all mortality experience of retired workers—but only by attained age and not with regard to duration of retirement. This experience is summarized in Table 1. For men there is very notable excess mortality at ages 65 and 66, but this differential gradually decreases for the older ages. This gives some indication of the higher mortality immediately after retirement. The effect thereof is diluted at the older ages as most of the experience is among continued lives rather than newly retired

ones. For women the same general tendency appears to be present although to a much smaller degree. The mortality of male retired workers at ages 75 and over very closely parallels population mortality, but for women the retired workers have 10–15% lower mortality at these ages and even at and shortly after age 65 the mortality is very close to that of the general population.

The old-age and survivors insurance data clearly indicate that con-

TABLE 1

RATIOS OF ACTUAL TO EXPECTED DEATHS AMONG RETIRED WORKERS[a] UNDER OLD-AGE AND SURVIVORS INSURANCE SYSTEM, 1950–52[b]

Age	Men	Women
65	136%	90%
66	145	107
67	128	99
68	121	98
69	116	94
70	115	90
71	111	90
72	107	87
73	106	85
74	105	85
75–79	99	82
80–84	98	85
85–89	101	90
90 and Over	103	100
All Ages	109	90

[a] Actually, includes all persons who claimed benefits even though some returned to work.
[b] Expected deaths based on U.S. 1950 White Lives Mortality Tables. Actual deaths: men 367,000; women 42,000.

siderably higher mortality than standard arises for individuals who have just retired, but this differential gradually reduces. This is particularly the case for men, although there is some indication of it also being present for women.

EXPERIENCE UNDER THE RAILROAD RETIREMENT PROGRAM

The railroad retirement program covering some 1½ million workers may be said to be a combination of an industry-wide private pension plan and a social insurance system since it contains elements of both.

In its actual operation, a very considerable amount of valuable mortality experience has been accumulated. In fact, it is the only large public retirement system for which good mortality data are available according to duration of retirement.

Table 2 compares the actuarial rates used in cost valuations for mortality of active workers and retired persons for ages 65 and 70. These have been tested against actual experience to a certain extent. According to these figures it is not expected that the mortality of these two groups will differ greatly at those ages where most retirements occur.

TABLE 2

TABULAR MALE MORTALITY RATES USED IN RAILROAD RETIREMENT SYSTEM[a]

(per thousand)

Age	Active Service	Age Retirements	Ratio of Active to Retired
65	30.2	30.2	100%
70	45.4	50.2	91

[a] Source: "Retirement Policies and the Railroad Retirement System," Part 1, Senate Report No. 6, 83rd Congress, 1st Session, pp. 341 and 357.

Table 3 compares the ratio of actual to expected deaths among age annuitants during a recent 3-year period. The characteristics of this plan are such that individuals may retire before age 65, with larger benefits if permanent and total disability is proved than if the retirement is for "age," and under certain circumstances retirement can be for "occupational" disability.

The mortality for age retirements at ages 60–64 is as much as 25% below the expected level during the early years of retirement although ultimately the mortality of this group approaches very close to that of the life table used as the basis of determining expected deaths. On the other hand, for those retiring at ages 65–69, actual mortality is appreciably higher than expected mortality—particularly in the first two years of retirement. This could be anticipated because those reaching age 65 who are in better health tend to continue at work and conversely those in poorer health retire. Those retiring exactly at age 65, relatively do not show as much excess mortality in the first few years of retirement as those retiring at ages 66–68. For those retiring at ages 70 and over, the mortality experience is quite close to that expected and

shows no significant fluctuation with duration of retirement. This might well be expected because age 70 is by employer practice virtually a compulsory age on most railroads. Accordingly this group is a good cross-section of persons of those ages—although perhaps somewhat healthier because they have been in employment up to that age.

TABLE 3

RATIO OF ACTUAL TO EXPECTED DEATHS AMONG RAILROAD RETIREMENT AGE ANNUITANTS, BY DURATION OF RETIREMENT, 1947–50[a]

Age at Retirement[b]	Duration of Retirement (Years) 0	1	2	3	4	5 and Over
65	112%	110%	99%	95%	89%	104%
66	135	118	115	99	105	114
67	123	114	109	116	124	111
68	141	132	110	109	105	115
69	111	110	95	107	97	105
60–64	74	87	75	78	96	101
65–69	121	114	104	101	99	107
70 and Over	103	93	103	101	104	104
All Ages	114	107	102	100	100	106

[a] Based on data furnished by Office of Director of Research, Railroad Retirement Board. Such data in summary form are contained in Table A-2, Annual Report of the Railroad Retirement Board for the Fiscal Year Ended June 30, 1952 (but shown there by attained age rather than age at retirement). Expected deaths based on 1944 Railway Annuitants Mortality Table, set back 1 year in age. Actual deaths: 25,545.

[b] Age last birthday.

Consideration of the railroad retirement data, in view of the specific provisions of that program, indicates quite clearly that the mortality of those who retire at and after age 65 is relatively high in the first few years of retirement. There is no conclusive evidence that this higher mortality is due to the act of retiring, but rather it seems probable that the retirements were to some extent caused by ill health which would have produced higher mortality anyhow.

EXPERIENCE UNDER CIVIL SERVICE RETIREMENT SYSTEM

The civil service retirement system covers some $1\frac{2}{3}$ million employees of the Federal Government and so is, in effect, a large self-administered pension plan. In general, depending upon length of serv-

ice, age retirement on full annuity can occur at ages 60 or 62. Prior to then, in certain cases, both disability and age retirement benefits are available, but the latter are in a reduced amount so that any disabled person would attempt to have his retirement adjudicated as due to disability.

Table 4 indicates the difference between the actuarial rates used in valuation of the system for the mortality of persons in active service and those who have retired on account of age. These two sets of figures should not be considered as reflecting the actual experience but rather give some indication of what is expected from an actuarial standpoint. For ages 60 to 70, the mortality of persons in active service is indicated to be some 30–50% lower than for persons who have retired on account of age.

TABLE 4

TABULAR MALE MORTALITY RATES USED IN CIVIL SERVICE RETIREMENT SYSTEM[a]

(per thousand)

Age	Active Service	Age Retirements	Ratio of Active to Retired
60	14.1	20.8	68%
65	17.6	30.9	57
70	21.5	46.5	46

[a] Source: Tables 27 and 31, 22nd Annual Report of the Board of Actuaries of the Civil Service Retirement and Disability Fund for the Fiscal Year Ended June 30, 1942.

Unfortunately, select data according to duration of retirement are not available for this system. Table 5, however, does show the ratio of actual to expected deaths by attained age for age retirements during a recent 3-year period. For men, the mortality experience under age 60 which is in respect to individuals who voluntarily retired on a reduced annuity and thus apparently could not prove disability was relatively low, just as was the case in the railroad retirement data. For attained ages 60–66, mortality is definitely higher than that according to the standard table, while at the older ages, the two tend to come together. Since this is an aggregate experience for all ages of retirement combined, it would be expected that this would occur at least after age 70, which is the compulsory retirement age. For women, the same general trends are evident except that there are greater fluctuations in the mortality ratios due to the smaller number of persons involved and

except that the mortality ratios for ages under 80 tend to show actual mortality well below that expected. This is not a significant factor in the experience as to the effect of retirement, but rather indicates that the standard table in use for women has too high mortality rates.

The experience under the civil service retirement program seems to

TABLE 5

RATIOS OF ACTUAL TO EXPECTED DEATHS AMONG CIVIL SERVICE RETIREMENT NON-DISABILITY ANNUITANTS, FISCAL YEARS 1950–52[a]

Age	Men	Women
Under 60	99%	40%
60	121	94
61	109	65
62	119	84
63	113	79
64	120	61
65	119	79
66	113	69
67	103	79
68	110	83
69	102	66
70–74	94	74
75–79	95	84
80–84	97	100
85–89	92	98
90 and Over	93	106
All Ages	99	81

[a] Based on data furnished by Retirement Section, U. S. Civil Service Commission. Expected deaths based on tabular rates shown in Table 4. Actual deaths: men—16,307; women—1,561.

confirm, in general, that of the railroad retirement system. Mortality is definitely lower than standard for those retiring at age retirements prior to the normal age and is definitely higher for those retiring at the normal age and a few years later. Again, this seems to indicate that the higher mortality shortly after retirement at or after the normal age is in considerable part due to the fact that ill health tended to cause retirement rather than vice versa.

EXPERIENCE UNDER PRIVATE PLANS

For many years insurance companies have collected experience under the group annuity plans which they sell primarily to commer-

cial and industrial concerns. In general, the annuities are payable beginning at age 65 regardless of whether the individual retires at that age, although in actual fact he may not receive the payment. Two subdivisions possible in the group annuity data are for "normal" retirements (generally payable from age 65 on) and "early" retirements, which in many—if not most—cases are disability retirements. As would be anticipated, the mortality under the "early" retirements is very high, especially at ages prior to 65, but subsequently tends to come closer to the mortality for the "normal" retirements (see Table 6). On the

TABLE 6

RATIO OF ACTUAL TO EXPECTED DEATHS AMONG MALE SERVICE PENSIONERS IN THREE SELF-ADMINISTERED PRIVATE PENSION PLANS[a]

Age	Group Annuity (1946–50)		Plan A[b] (1943–52)	Plan B[c] (1946–51)	Plan C[d] (1946–51)
	"Normal"	"Early"			
Under 55	[e]	312%	465%		
55–59	152%	248	280		
60–64	96	197	166		
65–69	102	149	116	86%	94%
70–74	112	136	118	107	102
75–79	119	137	123	128	117
80–84	137	113	131	147	120
85–89	124	[e]	120	117	129
90 and Over	127	[e]	128	[e]	110
All Ages	109	174	130	111	105

[a] Source: Report of Special Committee on Experience under Self-Administered Retirement Plans, *Transactions of the Society of Actuaries, 1953 Reports.* Expected deaths based on 1937 Standard Annuity Mortality Table. Actual deaths: Plan A—5,316; Plan B—613; Plan C—1,672.

[b] Group of public utilities covered under uniform plan.

[c] Electric utility company.

[d] Large company in electrical manufacturing industry.

[e] Insufficient data.

other hand, for the "normal" retirements, mortality shortly after age 65 tends to be somewhat low since payments generally begin automatically at age 65 and are thus made to quite healthy lives since a considerable number of disabled lives have already been eliminated as a result of the "early" retirements.

There has recently become available the first results of a continuing study by a committee of the Society of Actuaries in regard to the

mortality experience under self-administered retirement plans. As discussed previously, the resulting experience must be considered very carefully in view of the fact that the particular provisions in each plan will materially affect the results.

Table 6 compares the actual and expected deaths among male service pensioners under three privately administered pension plans. It should be noted that the mortality table used as a basis of the expected deaths is significantly too low at the extreme ages (beyond age 80) so that the mortality ratios developing tend to be artificially high.

Plan A has compulsory retirement at age 65 but has disability pensions prior to that age, which are included in this experience. Accordingly, as would be expected, there are very high mortality ratios prior to age 65, while those after age 65 tend to be somewhat above the group annuity "normal" retirement experience, at least between ages 65 and 75.

Both Plans B and C have compulsory retirement at age 65 and have separate disability benefits before age 65, the experience of which is not included here. As a result for ages 65 to 75, these plans show very low mortality since those entering the experience upon compulsory retirement at age 65 tend on the whole to be quite healthy lives. Certainly this latter experience would, of itself, not seem to give any indication that compulsory retirement produces high mortality.

SUMMARY

The preceding analyses of the mortality experience under various governmental and private pension programs indicate quite clearly that, in the absence of any special circumstances, the mortality of retired workers during the first year or two of retirement is considerably above the general level which otherwise might be expected but thereafter merges with such general level. It seems likely that this higher mortality in the early years of retirement arises from the fact those in poorer health are more apt to retire at or shortly after the minimum retirement age, while the healthier individuals continue at work.

An important factor to consider is that those retiring under a plan which does not have compulsory retirement generally tend to be the less healthy lives. On the other hand, in a plan providing for compulsory retirement at a particular age, those still in service at that age generally tend to be somewhat healthier than the normal population since they have recently been at work. Thus, it would be completely erroneous to contrast the mortality under a plan with compulsory retirement and that under a plan with voluntary retirement if there

were considered only pensioners. The results would seem to indicate lower mortality for the compulsory plan, which would not be a valid conclusion. It would really be necessary to contrast the mortality of pensioners under the compulsory retirement plan with that of both active employees and pensioners under the voluntary retirement plan. No such data were available to the author since usually mortality of active employees is not as closely studied as that for retired persons, particularly in governmentally administered plans. But if any progress is to be made in exploration of the subject of mortality after retirement, it will be necessary to obtain such data.[1]

The preceding discussion does not, however, mean that compulsory retirement might not have a serious effect on an individual's health and vitality, especially if he had not adjusted himself to the separation from employment. Unfortunately currently available data do not measure the effect of retirement on mortality after retirement. A priori reasoning would seem to indicate that compulsory retirement would certainly have some deleterious effect on mortality for some persons.

[1] The Department of Sociology and Anthropology of Cornell University is currently conducting a longitudinal study on the effect of retirement on mortality and morbidity, as between retirants and non-retirants under plans having different provisions as to retirement policy. For a description of this study see Milton L. Barron, Gordon Streib, and Edward A. Suchman, "Research on the Social Disorganization of Retirement," *American Sociological Review*, 17 (1952).

SAMPLING CONTROL OF LITERACY DATA*

S. S. Zarkovic
Federal Bureau of Statistics, Belgrade, Yugoslavia

An attempt is described to control the value of literacy data by the use of sampling methods. The reason for this research is the widely known unreliability of this sort of statistics in countries with a high rate of illiteracy. This research has been conducted as a part of the post-enumeration survey, taken in connection with the Yugoslav census of population as of March 31, 1954. The aim of the survey was the control of accuracy and value of different census results. The value of literacy data was checked on the sample of individuals by means of reading and writing tests. The results show (*i*) that literacy is a continuous variable, and (*ii*) the unreliable character of literacy statistics is connected with the difficulty of defining the limit between the different levels of literacy. Since these limits cannot be defined in the census of population, the best method to check the value of literacy data seems to be the use of sampling methods.

THE PROBLEM

Data on literacy are usually obtained in the census of population. Each person over a given age is asked about his ability to read and write.

What is the value of data provided in this way?

---

* The author wishes to express his indebtedness to Mr. S. Krasovec, formerly director of the Federal Statistical Office, and to Mr. M. Macura, director of the Serbian Statistical Office, who spent a lot of energy to make this research possible. The research described in this paper belongs to the new field in statistics that could be labeled "The problem of the value of statistical data." The most important work in connection with this problem has been done in the USA and India. So far obtained results and experiences can be found in the following papers: M. H. Hansen, W. N. Hurwitz, E. S. Marks, W. E. Mauldin: Response errors in surveys, *Journal of the American Statistical Association*, 46 (1951), 147–90, P. C. Mahalanobis: Recent experiments in statistical sampling in the Indian Statistical Institute, *Journal of the Royal Statistical Society*, 109; P. V. Sukhatme, G. R. Seth: Measurement of non-sampling errors, *Journal Indian Society Agriculture*, Vol. 4; P. V. Sukhatme: Measurement of observational errors in surveys, *Revue de l'Institute Internationale Statistique*, Vol. 20; M. H. Hansen, W. N. Hurwitz, L. Pritzker: The accuracy of census results, *American Sociological Review*, 1953; W. E. Deming: On errors in surveys, *American Sociological Review*, Vol. 9; E. S. Marks, W. P. Mauldin: Problems of response in enumerative surveys, *American Sociological Review*, Vol. 15; E. S. Marks, W. P. Mauldin, A. Nisselson: A case history in survey design; The post-enumeration survey of the 1950 census, *Journal of the American Statistical Association*, 48 (1953), 220–43; G. L. Palmer: Factors in the variability of response in enumerative studies, *Journal of the American Statistical Association*, 38 (1943), 143–52; S. S. Zarković: Completeness of enumeration (in Serbian), Federal Statistical Office, Belgrade, 1954; A. Gosh: Accuracy of family budget data with reference to period of recall, *Calcutta Stat. Assoc. Bul.*, Vol. 5; M. H. Hansen, W. N. Hurwitz, W. G. Madow: *Sample Survey Methods and Theory*, New York, Wiley, 1953; W. E. Deming: *Some Theory of Sampling*, New York, Wiley, 1950; S. S. Zarković: *Population Census Errors* (in Serbian), Federal Statistical Office, Belgrade, 1954.

The doubtful reliability of this sort of statistics is well known for it is clear that the answers are to a large extent the result of personal opinion of what literacy is. In *Population Census Methods*[1] one reads: "The meaning of the data on literacy and illiteracy obtained in a population census depends obviously to an important degree upon the extent of reading and writing ability that is assumed by the enumerators and respondents to be required for an affirmative answer."

To illustrate the unreliable character of these data we shall mention two examples.

In a European country with a very low percentage of illiterates, it was noticed after the mobilization during the last war that the percentage of those unable to read and write was far higher than was found during the preceding census. The situation among women was still worse. It is obvious the literacy situation, as depicted by the census statistics, is rather vague.

The next example is from Yugoslavia. The successive censuses gave the following percentages of illiterates:

1921—50.5
1931—44.6
1948—25.4
1953—24.9[2]

From this it appears that in the first 10 years illiteracy decreased by 6 per cent and in the next 17 years by 19 per cent in spite of the fact the schools were practically closed during the war. But in the last five years, during the regular work of all schools, with a very expanded system of education, including a great number of courses for teaching the alphabets and compulsory learning of reading and writing for those in the military service, illiteracy remained on the same level. There is something in this situation that deviates from a logical pattern.

On the basis of this figure for 1948, estimated with good reasons as being optimistic, while preparing for the census of 1953, we put in our program an investigation of the value of answers given by respondents on all census questions and consequently on the question of literacy as well.

This research proceeded in two directions. First, a sample of individuals was drawn immediately after the census, each of whom was requested by a specially trained inspector to answer again all census questions. Here the interest was in the stability of answers obtained in

[1] United Nations, 1949, p. 83.
[2] This figure is a sample estimate.

the census and in the extent of errors in them. The second research, also based on the sample of individuals, had as its aim to check, by means of tests, the degree of literacy of those who declared in the census that they were literate.

DATA ON THE SAMPLE[3]

To facilitate the organization of the census the whole country was divided in 118,999 enumeration districts (e.d.) with an average of 142 people, ranging in size from 0 to 300. For the purposes of sampling these e.d. were stratified in two strata, urban and rural. The urban stratum consisted of 29,805 e.d. and the rural one of 89,194. Before the beginning of the census only the total number of e.d. was known in each administrative unit and their distribution in strata. On this basis a random sample of 149 e.d. was drawn in the rural stratum and 100 in the urban one. Each e.d. was assigned equal probability.

In order to get individuals, subsampling was used. In the research on the stability of answers in the urban stratum the sampling fraction of 1:10 was applied to the total number of the enumerated people in an e.d. In the rural stratum the sampling fraction was 1:8. In this way a total of 1,682 people were investigated in the urban stratum and 2,470 in the rural one.

This gives the situation in Table 1.

TABLE I

SOME DATA ON THE SAMPLE

Stratum	Primary units (e.d.) in the sample		Secondary units (people)		Number of people in in the sample of secondary units
	Number	Percentage	Number of people in the sample of primary units	Percentage of the total	
Rural	149	0.167	19.818	0.163	1,682
Urban	100	0.336	16.785	0.347	2,470
Total	249	0.209	36,603	0.216	4,152

[3] Detailed information on this sample is given in S. S. Zarkovich: *Estimating Census Figures* (in Serbian), series "Studies and Analyses," No. 1, Federal St..tistical Office, Belgrade 1953.

The selection of secondary units was arrived at on the basis of census questionnaires, concentrated at that time in the commune office.

The task of inspectors was to find the people drawn into the sample and, using special control questions and available documents, to attempt to get the right answer on all census questions. In this work inspectors didn't know the answers given by respondents during the census.

These are data on the sample designed for the purpose of the general control of all the answers on census questions.

For the second research only those persons have been taken into account who: i) were, at the beginning of the census, 10 years of age and over, ii) put the answer "reads" and "writes" in the census questionnaire, iii) had an education of 4 years of elementary school or less. Those having more education were considered as definitely literate.

For this program the same sample of primary units was used as in Table 1. Since the definition of the population now was changed, the census returns were used again to select a new sample of secondary units. In this selection the sampling fraction of 1:6 in the rural stratum and 1:8 in the urban one was applied. So the sample consisted of 417 people in the urban stratum and 1,022 in the rural one.

In addition, another small sample was drawn of those who declared themselves illiterate. The purpose of this sample was to check whether this group of the population was homogeneously illiterate.

The individuals selected in this way came into a school where a test of their ability to read and to write was administered. Reading was investigated by means of 15 tests,[4] having each some printed phrases and three control questions that had to be answered by marking the right answer. Each right answer represented one point. The maximum number of points was 45. The testing of each group was limited to 10 minutes.

The ability to write was tested in a similar way. Our inspector dictated three phrases that had to be written in a limited time.

ANALYSIS OF ERRORS

When the field work was completed the control forms were matched against the census forms and the cases were defined as errors when the answers were not identical. So in connection with literacy 3.2 per cent wrong answers were found in the rural stratum and 2.8 per cent in the urban one. These percentages are calculated on the basis of total

[4] These tests were prepared in the Department of Psychology, University of Belgrade, by B. Stevanovich, N. Rot, Z. Vasich and M. Jovichich.

number of enumerated people in the e.d. All the present errors do not represent, however, the changed datum on literacy. Only 59.0 per cent (36 people) out of total number of errors in the urban stratum and 65.4 (53 people) in the rural stratum represent the changes in the answer previously given. The remainder of errors covers the omissions of answers for the people of 10 years and over or giving answers for children under 10 years.

The next problem is as follows: is there any tendency among individuals to declare themselves literate when they are not so or to declare themselves illiterate when really literate? Letting the first type of tendency be represented by + and the second by −, we found the following distribution of changed data:

Stratum	Males		Females	
	+	−	+	−
Urban	3	3	16	14
Rural	5	8	15	25

These figures do not agree with what is rather general opinion, viz. the existence of a marked tendency among individuals to declare themselves literate even if not so. If there is a place here for any tendency it should be stated in just the opposite way (in accordance with the results for the rural stratum).

This conclusion might appear somewhat hazardous since both minuses and pluses may conceivably represent errors in the second report. But we think it is safe within the best possibilities of the checking procedure in this field. Each changed answer was subject to a special investigation.[5] So only some intermediate cases (*vide infra*) might represent the problem.

SOURCE OF ERRORS

Now we face the very important practical question of the source of these errors.

The information given by the respondents with the wrong census answers shows that two main sources of errors exist. In the first group the extreme cases are included, namely individuals being absolutely literate or illiterate. In the second group the intermediate cases are involved.

[5] The description of the checking procedure is giver in S. S. Zarkovich: *Population Census Errors.* Belgrade 1954.

By the definition of the group, in the first case the answer on the question of literacy is known. If someone never learned reading and writing or if a person had a college education it is clear what the answer should be. But the errors still appear. For absolute illiterates we found answers "literate" and vice versa.

In our census the source of these errors is the system of enumeration. In our system the questionnaires were distributed one day before the beginning of the census and collected the day afterwards. Meanwhile everybody was supposed to fill answers personally (if literate). For children and illiterate people the giving of answers was the duty of parents or some other member of the family. The enumerator had to check and correct data given or to put down answers in the case when no one was able to do it (in villages).

Now, if a member of a family fills the questionnaires for the others he may not always be well informed on what the answer should be. It particularly holds for the people on the lower cultural level where no attention is paid to literacy. If the enumerators do this job on the basis of information given by the head of the family the errors appear in the same way. It would be the best if the enumerators had a separate talk with each respondent. In this case the number of errors would probably be less serious.

Consequently, these errors can be influenced primarily by changing the system of enumeration (if there is any possibility to do so). The recommendation in *Population Census Methods* by which the "general adoption of the criterion . . . ability to read and write a simple message in any language, would help to improve the comparability and meaningfulness of census statistics on this subject" does not seem to be useful.

In the group of intermediate cases the errors appear because the person in the low level of literacy declares himself literate and vice versa. Here the respondents *don't know* what their answers should be like. To what degree should the ability to read and write be developed to entitle either of two possible answers? The problem is the limit that divides literacy from illiteracy. Considering the fact this limit can only be defined in terms of some units it is obvious that no system of enumeration is likely to change the frequency of errors. It also seems that the above recommendation in *Population Census Methods* couldn't be expected to be helpful. We found a lot of people able to distinguish any letter and read any word but the reading represented a tremendous effort for them in which they used 20 times more time than a man with a university education. From the point of view of the "ability to read

and write a simple message" they are literate but from any practical point of view they are illiterate. In general, such an individual does not use at all his ability to read and write because this is for him as painful a job as any other in which great physical efforts are concerned. The limit of such a literacy is the illiteracy.

MEASURING DEGREE OF LITERACY

Literacy is a continuous variable with illiteracy and complete literacy on its extremes. Any intermediate value brings about the question whether it should be called literacy or not. The difficulties appear because of the very nature of literacy as a census characteristic. If it is desired to have a more precise insight in what the literacy of the "literate" people really means, there is no other way—as it seems to me—than to draw a sample of "literate" individuals, apply some measuring and, on the basis of the results received, estimate the percentage of the people on different levels of literacy.

This was the aim of our second research. In connection with this the following had to be done:

i) prepare tests for measuring,

ii) define the limits between different classes of literacy in terms of units of these tests,

iii) apply these tests and calculate the percentage of people in each class.

The results received are shown here in Tables 2 and 3. In the stubs of these tables is the reading score and in the headings, the writing score.

These figures show the existence of correlation between the ability to read and the ability to write. Then, one sees that there is a percentage of people with a pretty high score in reading and a low one in writing (the first two columns). They also show that among the people with moderate or high scores in writing there is a percentage of those (the first row) who did not understand what they read. These have no points in reading.

At the same time these data show that even in the class of the literate people (because in this research only those have been included who declared themselves "literate" in the census) there is a number of those who didn't get any points in either reading or writing (first column, first row). These are illiterate. Most of these cases were separately investigated and their illiteracy was proved. It was found that their presence in the class of literate people was due to the system of enumeration: those giving answers for them put them in this group.

On the basis of these tables, the possibility of estimating the real

TABLE 2

DISTRIBUTION OF SCORES IN READING AND WRITING
(Urban stratum)

Reading Score	Writing score										Total
	0	1	2	3	4	5	6	7	8	9	
0	5	–	1	–	–	–	–	–	–	–	6
1–5	3	1	1	3	2	5	4	1	–	2	22
6–10	4	–	2	4	6	5	6	4	1	–	32
11–15	–	1	3	2	9	7	15	4	10	7	58
16–20	2	1	1	2	1	8	21	7	9	12	64
21–25	1	1	–	–	2	5	13	9	23	21	75
26–30	–	–	–	2	–	2	5	8	21	21	59
31–35	–	–	–	–	1	–	1	3	11	18	34
36–40	–	–	–	–	–	1	1	3	12	13	30
41–45	–	–	–	–	–	–	1	2	11	23	37
Total	15	4	8	13	21	33	67	41	98	117	417

TABLE 3

DISTRIBUTION OF SCORES IN READING AND WRITING
(Rural stratum)*

Reading score	Writing score										Total
	0	1	2	3	4	5	6	7	8	9	
0	17	1	2	1	1	3	2	–	1	–	28
1–5	10	7	11	6	3	6	14	3	–	–	60
6–10	4	8	13	11	15	33	18	9	10	5	126
11–15	5	4	4	4	17	32	37	19	18	10	150
16–20	3	4	2	3	11	20	24	23	38	18	146
21–25	1	2	1	1	3	12	31	24	42	29	146
26–30	–	–	–	2	5	7	24	24	40	34	136
31–35	–	1	–	1	3	4	7	9	34	21	80
36–40	–	–	–	–	–	3	6	7	30	18	64
41–45	–	–	–	1	2	1	8	10	23	41	86
Total	40	27	33	30	60	121	171	128	236	176	1022

number of literates by combining the scores both in reading and writing is obvious. Because of the small sample, our problem was the trichotomous classification. Experiments to give the respective values of limits are now in process.

But another illustration of the limits can be given here. Taking into consideration only the errors in reading, it was found that the man with a university education takes on the average 6 minutes to get through all tests. In doing so he generally makes no errors and gets 45 points. In other words, one point takes approximately 8 seconds. In the same way the score of 5 points in 10 minutes means an average of 120 seconds per point or 15 times slower reading than a man with university education. Adopting now the criterion that the score of 15 or more times slower reading than our standard, i.e. 0–5 points, represents the class "illiteracy," then that 5–15 points or 15–5 times slower reading defines the "middle literacy" and 15 points and over the class "literacy," our data give the following distribution:

Degree of literacy	Stratum	
	Rural	Urban
Illiteracy	6.7	8.6
Middle literacy	21.6	27.0
Literacy	71.7	64.4

To complete these data Figure 1 is also given. On the x-axis is the score in reading and on the y-axis the percentage of persons having reached the respective score. This figure is characteristic for the problem of literacy.

Consequently, the results of such research can be used to enlarge the knowledge of what is behind the general term "literate" people.

CONCLUSIONS

Data on literacy, as obtained by the census, probably are not very reliable in any country, although the degree of their value may vary considerably in connection with specific conditions. The possibility of influencing their reliability by means of better definition is also doubtful.

There are two main problems in this field: i) the value of the answers of those absolutely literate and absolutely illiterate and ii) the quality of answers of intermediate cases, i.e. of the people who are neither completely literate nor illiterate. In the first case the right answer depends mostly upon the system of enumeration and in the second one upon the lack of a criterion as to the level of ability to write and read at which literacy begins. To control the meaningfulness of data belonging to this second class, an experiment of the described sort is very useful.

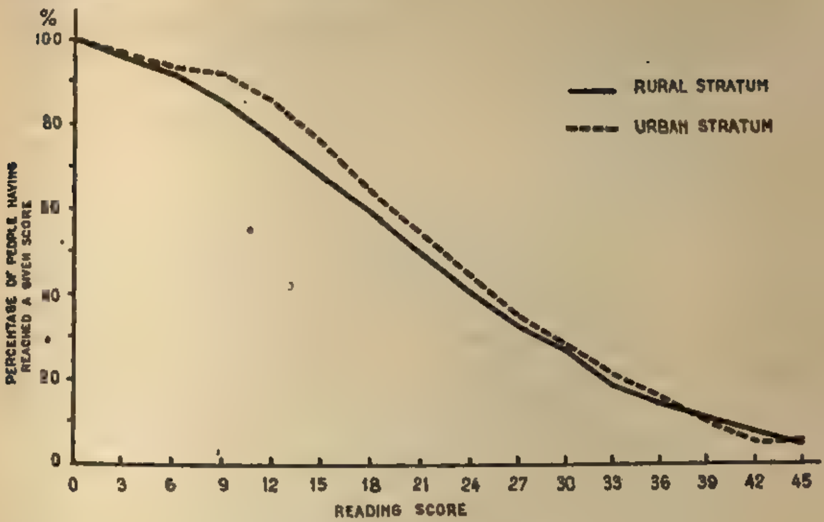


FIG. 1. Decreasing degree of literacy.

Some may agree to the usefulness of such a control but doubt might arise as to the possibility of carrying through a similar large-scale research. The problem may especially arise in connection with the willingness of the people to be tested.

Perhaps some words on our experience will be useful here. To carry on this experiment we used 250 young employees of the statistical office who had been trained two to three hours a day during less than two weeks. They were charged with the whole field work in the application of sampling methods in connection with this census. Most of them had never had any contact with psychology and education, but in spite of that their technique of experimentation was considered by experts as very satisfactory.

On the other hand, we did not have any difficulties with people. Before these experiments started, our inspectors contacted respondents in connection with the control of the completeness of enumeration and the control of all answers in the census questionnaire. At this occasion they also had a contact with those people selected for this experiment and explained to them the purpose and the sense of this work. The result was that out of 1439 primarily selected persons only 14 didn't come to the testing place. They have been replaced by the others selected at random as well.

If this experience has any meaning for other countries, I should conclude that the sampling control of the literacy data does not raise serious difficulties.

RESPONSE ERRORS IN ESTIMATING THE VALUE OF HOMES*

Leslie Kish *and* John B. Lansing
Survey Research Center, University of Michigan

In the 1950 Survey of Consumer Finances home owners were asked to estimate the market value of their houses. Estimates for these same homes were later made by professional appraisers. These two estimates for each of 568 homes comprise the data analyzed here. The proportion of discrepancies between the two estimates is great: only 37 per cent of the estimates by respondents are within plus or minus 10 per cent of the appraisers' estimates. However, the errors tend to be offsetting, and in none of the ten price classes used is the difference in the relative frequencies for owners and appraisers statistically significant. Similarly, although the root-mean-square difference between the two measurements is high (an average of $3,100), the mean of the respondents' estimates is only $350 higher than the mean of $9,200 for the appraisers' estimates. The amount of variability is found to be rather similar for several sub-populations. However, for houses worth over $10,000 the mean-square difference between the measurements is found to increase with the value of the home. In the Appendix a model is developed for the statistical investigation of the data.

INTRODUCTION

Knowledge of the over-all financial position of consumers has been a primary objective of the Survey of Consumer Finances conducted annually since 1945 by the Board of Governors of the Federal Reserve System in cooperation with the Survey Research Center of the University of Michigan.

More than half of American families live in their own homes, and for the vast majority of these families, that home is their most valuable single asset. To be complete, then, any analysis of the financial position of consumers must cover this asset.

In the 1950 Survey of Consumer Finances, respondents were asked to give their idea of what their house was worth.[1] The answers they

* The authors are indebted to Clarke L. Fauver of the staff of the Federal Reserve Board, who initiated the research reported here while he was on the staff of the Board's division of Research and Statistics. They are also indebted to the American Institute of Real Estate Appraisers, the Federal Housing Administration, and the Society of Residential Appraisers for their participation in the field work.

[1] For a discussion of the methods used in this survey see G. Katona, L. Kish, J. B. Lansing and J. K. Dent, "Methods of the Surveys of Consumer Finances," *Federal Reserve Bulletin*, 36 (1950), pp. 795–809.

gave have been tabulated and on the basis of those replies tables were published in the *Federal Reserve Bulletin*[2] showing distributions in class intervals of owners' estimates of the current value of their homes. The published distributions are for all owners, for owners having different incomes, owners with different occupations, and owners living in towns and cities of different sizes. This is important basic information for the student of housing economics.

The question naturally arises, how reliable are these data? How much does the average householder know about the going market price for his house? Assuming for the moment that he does know, is his answer to the interviewer's question likely to be seriously biased? Are recent buyers of homes more informed about current market conditions than owners who may have bought many years earlier?

It was in an effort to answer some of these questions that a special attempt was made to evaluate the responses given to questions concerning house values in the 1950 Survey of Consumer Finances. Respondents who reported they owned their own homes were asked in January and February 1950 whether they had purchased their homes in 1949 or in some earlier year. Those who had purchased before 1949 were asked: "Could you tell me what the present value of this house is? I mean about what would it bring if you sold it today?" (A similar question was asked in the 1950 Census of Housing.) Those who had purchased their homes during the year 1949 were asked: "How much did the house and lot cost?"

Subsequent to the completion of interviewing, it was decided to check the estimates of respondents by obtaining estimates from qualified residential appraisers. Through the cooperation of the American Institute of Real Estate Appraisers, the Federal Housing Administration, and the Society of Residential Appraisers, arrangements were made to have professional appraisers visit a substantial number of the properties. The appraisers were not required to obtain access to the property; they were asked to look at it from the outside and to estimate its value in the light of their experience and familiarity with local real estate conditions.

From the sample of home owners found in the yearly survey a subsample was selected, including respondents who failed to answer the questions about home-ownership, but not including any potential respondents who had not been interviewed during the regular survey.

[2] J. A. Frechtling, J. H. Lorie and Irving Schweiger, "1950 Survey of Consumer Finances, Part V, The Distribution of Assets, Liabilities, and Net Worth of Consumers, Early 1950," *Federal Reserve Bulletin*, 36 (1950), pp. 1595–97.

(In the subselection a higher probability of selection was given to the more extreme house values.) The sample was distributed roughly evenly among the three participating organizations. The response rate in the follow-up study was 89 per cent. This high response rate, a result of the excellent cooperation on the part of these professional groups, made possible the analysis which follows.

The number of homes used in the first stage of the analysis is the 637 for which forms were returned. In 30 of the 637 cases the value of the property was not indicated on the completed form. In an additional 39 cases the respondent failed to give a usable answer to the question in the original survey. Hence there are 568 homes for which two estimates of value are available. (In calculating the response rate of 89 per cent mentioned above these 69 cases were treated as responses since some useful information is available about them. If the 69 were classified as non-responses, the response rate would be 79 per cent.)

Essentially, the analysis was divided into two stages. The first stage involved a simple comparison of the frequency distributions and cross tabulations obtained by the original survey and by the follow-up study. The second stage involved the statement of a mathematical model of the response error, and estimates of the terms of the basic equation of this model. Although the conclusions drawn in the second stage are described in the main body of this article, the model itself appears in the Appendix.

COMPARISONS OF CELL FREQUENCIES

The first step in the analysis was to compare the frequency distribution obtained from the survey of owners with that from the survey of appraisers. The results of the comparison appear as the first two colums of Table I. Columns (3) and (4) show cumulative totals for columns (1) and (2), respectively.

The fifth column of Table I shows the distribution of appraisers' estimates for the 39 cases which were "Not Ascertained" in the survey. On seeing the "NA's" in any table one is led to wonder about their effect on the entire distribution. There is a mere suggestion of a concentration of several homes with very low values among these 39 cases. But anyone who assumed that the 39 cases should be distributed proportionally would not have been led far astray.

In column (6) the differences between the entries of columns (1) and (2) are given. These differences are subject to sampling variability. If we sent out both interviewers and appraisers to repeated samples of 600 cases under identical conditions, we would expect that the bracket

distributions sometimes would show closer agreement than (1) and (2), and sometimes wider disagreement. The model in the Appendix permits us to estimate the probability that the proportions in any pair of cells will agree within a given range. That is, we can estimate how the differences shown in column (6) would fluctuate if the present study were repeated many times. The measures of this fluctuation, estimated

TABLE I

FREQUENCY DISTRIBUTIONS OF THE VALUE OF OWNER-OCCUPIED HOMES BASED ON ESTIMATES REPORTED BY OWNERS AND APPRAISERS (UNCORRECTED)[a]

(percentage distribution of homes)

Value of Home	Respondents' Estimates (1)	Appraisers' Estimates (2)	Cumulative Total of Respondents' Estimates (3)	Cumulative Total of Appraisers' Estimates (4)	Appraisers' Estimates Where Respondents' Were Not Ascertained (5)	Difference Between Proportions (1)–(2) (6)	Standard Error of the Difference in (6) (7)
Under $2,500	2.9	2.3	2.9	2.3	14	+0.6%	0.7%
$2,500– 4,999	13.1	13.7	16.0	16.0	14	−0.6%	1.4%
$5,000– 7,499	19.6	19.3	35.6	35.3	20	+0.3%	1.9%
$7,500– 9,999	21.5	24.3	57.1	59.6	18	−2.8%	1.9%
$10,000–12,499	19.1	16.8	76.2	76.4	7	+2.3%	1.8%
$12,500–14,999	6.5	8.8	82.7	85.2	10	−2.3%	1.2%
$15,000–19,999	7.2	6.3	89.9	91.5	3	+0.9%	1.1%
$20,000–29,999	2.8	2.2	92.7	93.7	3	+0.6%	0.7%
$30,000 and over	1.5	1.4	94.2	95.1	3	+0.1%	0.4%
Value not ascertained	5.6	4.7	99.8	99.8	8	+0.9%	1.2%
Total	99.8[b]	99.8[b]			100		
Number of homes	637	637	637	637	39		

[a] These "uncorrected" distributions contain clerical errors which were discovered and corrected in the course of comparing the data from respondents and appraisers. Later tables are based on corrected data except as indicated.

[b] Detail does not add to 100.0% owing to rounding.

from the data of this study, are presented in column (7) in terms of the standard errors of the differences. We may illustrate the interpretation of these columns as follows: the discrepancy between the proportion of homes placed in the bracket $2,500–4,999 by respondents and appraisers was 0.6 per cent in the present study; if the study were repeated many times, this difference would be less than 1.4 per cent in two studies out of three in the long run, and it would be less than 2.8 per cent in 19 studies out of 20.

The two distributions in (1) and (2) convey the same general impression about the proportion of owner-occupied homes of different values. The same would be true of other similar distributions from replications of the present study in view of the relatively small size of the errors shown in (7). We make this judgment (and ask the reader to do likewise) within the general framework of the errors and requirements of surveys of this kind and size. It would be fruitless for us to raise here the question: for what kind of decisions are our results "reliable enough"? Our investigations do provide assurances against the existence in the procedures of large response errors. "Large" here is taken in the context of the actual sizes of the sample and of the sampling errors—but we must neglect the question of the relative cost of reducing the response error.

Although we find no reliable evidence of a net bias in any price class, it is possible (and even probable) that a large enough sample would uncover biases which escape detection in this sample. We have shown only that the differences between columns 1 and 2 *could* be the result of random response variation.

The second step in the analysis was to examine the discrepancies between the estimates of respondents and appraisers. The similarity between the first two columns in Table I could be the result either of few errors or of many off-setting errors. Table II compares the classification of the homes by respondents and appraisers. A sum of the proportions in the cells along the diagonal indicates that 43 per cent of the homes that were included in a given bracket by the respondents were also placed in that bracket by the appraisers. Errors were, in fact, frequent, but generally off-setting.

On close examination some of the differences shown in Table II seemed out of all reason—how could *any* house valued by a respondent at under $2,500 be valued by an appraiser at over $15,000? This question raises the possibility of errors in the survey process made by others than respondents and appraisers. The information in Table II was used to guide a special search for errors. All cases where the two estimates were in disagreement by more than one "bracket" (coded class of house value) up to a value of $15,000, and above that value all cases not in the same "bracket," were selected for study.

This search involved a comparison of the original interview, the appraiser's report, and the card on which the data had been punched. Such a search is unlikely to turn up errors by interviewers in recording the answers given by respondents, but it should disclose any errors in

coding. An examination of 109 cases yielded 17 errors, all but two of them clerical errors by coders. Of the 17, four involved only errors in the conversion of a dollar amount (entered correctly) to a bracket (entered wrongly). There were 11 clerical errors made in coding the respondents' estimates, and two errors were made by interviewers. Ten of these 11 errors involved entries of one-tenth of the proper amount owing to the omission of a zero; in the one case, $11,000 was read as $77,000. In addition to these errors two exceptional cases were

TABLE II

RELATION BETWEEN APPRAISER'S ESTIMATE AND RESPONDENT'S ESTIMATE (UNCORRECTED)[a]

(percentage distribution of homes)

Appraiser's Estimate	Respondent's Estimate										
	Under $2,500	$2,500 –4,999	$5,000 –7,499	$7,500 –9,999	$10,000 –12,499	$12,500 –14,999	$15,000 –19,999	$20,000 –29,999	$30,000 & Over	Value not ascertained	Total
Under $2,500	1.0	0.4	0.2[b]							0.7	2.3
$2,500– 4,999	0.7	7.2	3.8	1.1	0.2					0.7	13.7
$5,000– 7,499		3.4	8.3	4.8	1.7					1.1	19.3
$7,500– 9,999	$\boxed{0.2}$	0.5	5.0	11.1	5.7	0.2	0.6			1.0	24.3
$10,000–12,499	$\boxed{0.4}$		0.3	3.8	7.8	2.4	1.5	0.2		0.4	16.8
$12,500–14,999	$\boxed{0.2}$		0.7	0.1	2.3	3.0	1.2	0.6	$\boxed{0.1}$	0.6	8.8
$15,000–19,999	$\boxed{0.2}$				0.9	0.7	3.2	1.0	0.1	0.3	6.3
$20,000–29,999		$\boxed{0.2}$					0.6	0.8	0.4	0.2	2.2
$30,000 and over				0.2[b]				0.2	0.8	0.2	1.4
Value not ascertained	0.2	1.4	1.3	0.4	0.5	0.2	0.1		0.1	0.5	4.7
Total	2.9	13.1	19.6	21.5	19.1	6.5	7.2	2.8	1.5	5.6	99.8
Number of cases[c]	23	80	108	120	105	39	50	47	26	39	637

[a] For the difference between corrected and uncorrected data, see the discussion in the text. The principal effect of the corrections on this table is to empty the cells indicated by boxes, distributing the entries among other occupied cells. The table reads as follows: 1.0% of all houses in the sample were valued at under $2,500 by the respondent and also by the appraiser; 0.4% of all houses were valued at $2,500–$4,999 by the respondent, but at under $2,500 by the appraiser; etc.

[b] These two cells contain one case apiece. They are the exceptional cases noted in the text.

[c] Because there were three different weights used, the percentages are not simple ratios of the total of 637.

noted. In one, it seemed clear that the appraiser had included only part of the property which the respondent had in mind. In the other, the appraiser based his estimate on the commercial value of the property, while the respondent based his on the value for residential purposes.

The effect of these errors on the entries of a few cells in Table II are

shown: certain cells which are emptied by the corrections or which contain only exceptional cases, have been indicated by being enclosed in "boxes." All of the most extreme discrepancies in Table II disappear but the marginal distributions are little changed by the corrections. It is interesting to note that the lowest class was composed in large part of errors.[8]

The comparison in Table II is supplemented by another approach in Table III: this presents the distribution of each respondent's estimate divided by the appraiser's estimate on his home. This division was carried out for 568 homes. The respondents' estimates were within plus or minus 10 per cent of the appraisers' in 37 per cent of the cases. On the other hand, the discrepancy was more than plus or minus 30 per cent for 24 per cent. Of these 24 per cent, 18 per cent represent overestimates by respondents, suggesting a tendency for owners to overvalue their homes. This possibility can be better evaluated by comparing the means of the two distributions (after correction of the clerical errors).

TABLE III

FREQUENCY DISTRIBUTION OF RESPONDENT'S ESTIMATE DIVIDED BY APPRAISER'S (IN BRACKETS)[a]

Respondent's Estimate Divided by Appraiser's	Proportion of Homes
Under 70%	6
70– 89%	20
90–109%	37
110–129%	19
130–149%	9
150% and over	9
Total	100
Number of homes	568

[a] Uncorrected for the clerical errors discovered only after comparison of the two estimates.

[8] This finding and the $77,000 mistake have obvious implications for checking procedures. It should be noted that the checking procedure used in processing the 1950 Survey of Consumer Finances varied according to the nature and extent of the projected analysis of the data. The data on value of houses received the minimum amount of checking. For the type of distribution actually published in the *Federal Reserve Bulletin* these clerical errors were of little importance. The clerical errors do have a large effect on the errors of the estimated *mean* value; however, the mean was neither intended nor submitted for publication.

COMPARISONS OF THE MEANS OF THE TWO DISTRIBUTIONS[4]

The difference between the means obtained by the two methods of measurements is $9,560 − $9,210 = $350. That is: the mean of $9,560 obtained from the responses of the home-owners seems to include a bias of $350 (if we accept the appraisers' values as "true"). This bias is in the direction one would expect. The standard error of the difference was calculated (by a formula proper to the complexities of the sample design) to be $170. Hence there appears to be a tendency (statistically significant) for the home-owners to set higher values on their homes than do the professional appraisers. This tendency is small compared to the value of the home—about 4 per cent of the latter.

This net average bias may appear small also in comparison with the large discrepancies found in the two values obtained for individual homes. The mean square difference of the two measurements was estimated as 9,580,000 compared to the estimate of the squared bias of 100,000 (see equation 15 in Appendix). This result is consistent with the findings presented in Tables I and II which show also large discrepancies in individual estimates but small differences in the overall distribution of the two measurements.

The relative importance of a bias depends on the size of the survey to be taken. The sample mean of a simple random sample of n interviews with respondents may be expected to be subject to a total root-mean-square error of $\sqrt{[V(r)/n]+\overline{D}^2}$ where the first term under the radical represents the total variability of the estimates from the survey of respondents about their own mean and the second, the square of the bias.[5] As the size of the sample increases the first term will decrease but the second will remain constant.

We may use the sample estimates obtained in our investigation to examine the effect of the bias on the total error. For $V(r)$ we have the estimate $v(r) = 32{,}650{,}000$; and for $\overline{D}^2$ we have the estimate $\bar{d}^2 = 100{,}000$ (see equation 15 in Appendix). Now let us take the value of $\sqrt{[\bar{V}(r)/n]+\overline{D}^2}$ for three different sample sizes, and under the two assumptions: that $\overline{D}^2 = 100{,}000$ and that $\overline{D}^2 = 0$.

---

[4] We include this analysis because it may be of general interest. We repeat: the mean home value was not sought nor published in the original survey.

[5] The term $V(r)/n$ represents the variance of the sample mean as it is usually calculated, but it actually includes both the error resulting because not every member of the universe was interviewed—the sampling error proper—and any uncorrelated random response error which may be present in the methods used, such as random clerical errors (see equation 7 in Appendix). The *net* average of the response errors will be reflected in the squared bias term, $\overline{D}^2$. This expression shows that it is possible to increase the accuracy of the estimate of a mean from a simple random sample in one of three ways: by increasing the number of interviews (increasing n); by reducing the variability of the estimates, $V(r)$, (by reducing some of the errors of response); or by reducing the size of the bias $\overline{D}^2$ (for example, by more careful training of interviewers). The practical problem in the administration of surveys is to allocate resources among the three in such a way as to minimise the total error for a given outlay.

The total (root-mean-square) error of the mean house value under six different conditions would then be as follows:

Value of D^2	Formula	Value if Sample Size (n) is		
		100	1,000	10,000
(1) Total error if $D^2=100{,}000$	$\sqrt{\dfrac{32{,}650{,}000}{n}+100{,}000}$	$650	$360	$320
(2) Total error if $D^2=0$	$\sqrt{\dfrac{32{,}650{,}000}{n}}$	$570	$180	$60

Note that the total error for a sample of 100 is not greatly increased by the bias term, but, for a sample of 1,000, the effect of the bias term is large, while for a sample of 10,000 it is overwhelming. Where the facts are similar to those found in this investigation, an improvement in the accuracy of the estimate for surveys of a few hundred cases can probably be obtained most easily by increasing the size of the sample. For surveys of several thousand cases, however, it may be more efficient to allocate funds for a search for sources of bias and for development of techniques for reducing the bias than to allocate funds for an increase of the size of the sample.[6]

We have investigated the possibility that the discrepancy between the two measurements might prove to be a function of the value of the house. One can imagine, for example, that respondents might tend to overvalue low priced homes and to undervalue high priced homes. We divided the homes into groups based on the appraisers' value and estimated the mean value of the houses in each group, first on the basis of the respondents' values for the houses and then on the basis of the appraisers'. The difference between these two means has been plotted in Chart 1. (See solid line "A.") This graph indicates that the respondents tend to overvalue homes priced below about $12,000. For homes priced above that amount no clear tendency to under- or over-valuation appears. We feel that the discrepancy below $12,000 may be explained in part, but only in part, by errors in the estimates made by appraisers. Any such errors also would tend to give this graph a general trend downward to the right.[7]

[6] From the data of Table IV of the Appendix it seems that the contributions of the bias term are less for estimates of the proportions of houses falling into designated class intervals.

[7] For example, suppose the respondent estimated the value of this house correctly at $11,000. The point on the chart to correspond would be $X=\$11{,}000$, $Y=0$, if the appraiser made the same estimate. However, if the appraiser made an estimate of $10,000, the corresponding point would be $X=\$10{,}000$, $Y=\$1{,}000$, which is above and to the left of the original point. An error by the appraiser in the other direction would lead to a point below and to the right of the original point.

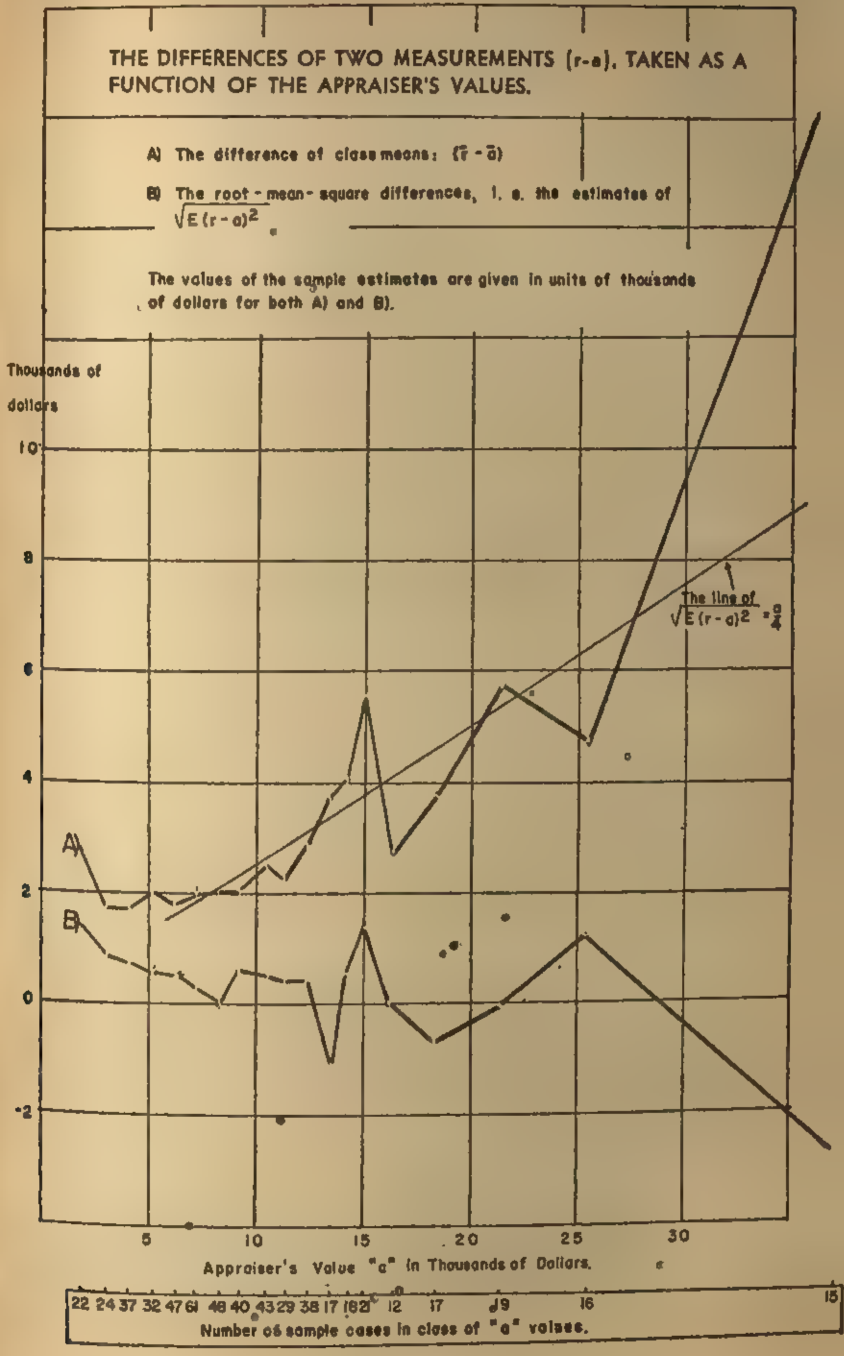

THE DIFFERENCES OF TWO MEASUREMENTS (r-a), TAKEN AS A FUNCTION OF THE APPRAISER'S VALUES.
A) The difference of class means: $(\bar{r}-\bar{a})$
B) The root-mean-square differences, i. e. the estimates of $\sqrt{E(r-a)^2}$.
The values of the sample estimates are given in units of thousands of dollars for both A) and B).
Thousands of dollars
10
8
6
4
2
0
-2
A)
B)
The line of $\sqrt{E(r-a)^2}=\frac{a}{4}$
5
10
15
20
25
30
Appraiser's Value "a" in Thousands of Dollars.
22 24 37 32 47 61 48 40 43 29 38 17 182 12 17 19 16 15
Number of sample cases in class of "a" values.

CHART 1

THE ROOT-MEAN-SQUARE DIFFERENCE

As a measure of the average individual discrepancy between respondents and appraisers we use the root-mean-square difference, that is, the square root of the mean of the squared deviations between the pairs of estimates. We estimate this quantity at $3,100 for the sample as a whole. In other words if we assume that the appraiser's estimate is the true value, the respondents are in error by an average of $3,100 in their estimates. (From equation 15 in Appendix.) Actually there is no doubt that the appraisers also made errors, and the average discrepancy between the respondents' estimates and "true value" of the property would be less than $3,100.

How does the average discrepancy vary with the value of the home? Is the discrepancy a constant amount, or a constant proportion of the house, or some other function? On Chart I are plotted the root-mean-square differences—the r.m.s.(d) values—for each class of appraised values; the width of the intervals is $1,000 except at the ends, where classes were combined to obtain larger cells. (See the solid line "B.") For values below $10,000 the r.m.s.(d) appears to be constant around $2,000. For values above $10,000 it is considerably more variable and larger; and, it appears to be proportional to the estimated value of the home. The line which represents a root-mean-square difference of one-fourth of the appraised value is drawn in. It appears to the eye to fit the distribution above $10,000 fairly well. In other words in our data the expected absolute value of the difference between the respondents' and the appraisers' estimates is about $2,000 for a house worth less than $10,000; while, for a house worth over $10,000, the expected value of the difference is one-fourth of the appraisers' estimate. For a $16,000 house, one would predict a respondent would differ from an appraiser by $4,000; for a $20,000 house, one would predict $5,000, and so forth.

ANALYSIS OF SOME SUB-GROUPS

One aim of our investigation was to discover some of the variables which might be associated with response errors. For three cross-tabulations comparisons were made of the ratio of respondents' to appraisers' values. An attempt was made in the original survey to isolate those cases where the respondent seemed uncertain of his estimate. If this attempt were successful, it was thought that it might be possible to develop methods of analysis that would place more weight on the more reliable cases. The procedure tried was to instruct the interviewers as follows:

Since some respondents have a very clear idea of the value of their house,

> based on such things as what the house next door just sold for, while others have only very vague notions, we have left space after question 31 in which you should note down any information he may give you about how he arrived at his estimate of the value of the house. Our objective is to distinguish between cases where we have the kind of accurate estimate we would prefer and cases where we have only vague information. In any case be sure to record the dollar value of the house.

The coders were then instructed to study the answer as recorded by the interviewer and attempt to assign a rating according to how sure the respondent seemed to be of his answer. This rating proved very difficult to make; coders disagreed frequently as to the proper point on the scale at which to place an answer. The relevant data (not given here) show that the assigned rating of the appearance of reliability had no validity: the errors were about equally large in the various classes of assigned reliability.

Secondly, occupation of the head of the family owning the house was selected as a measure of socio-economic status, on the hypothesis that people of higher status might be better informed. Thirdly, the population of the place (city) of residence of the respondent was selected on the hypothesis that knowledge of real estate values would be different in communities of different sizes. None of these hypotheses were substantiated; no sizeable differences were noted.

For four subgroups of the sample we calculated separately the estimates of our basic error equation (8). There exist *a priori* reasons why the accuracy of the estimates in each of these groups might turn out to be different than in the entire sample. The calculated equations are in the Appendix; here we shall summarize the results, using the root-mean-square difference—r.m.s.(d)—as the measure of accuracy. The conclusions we draw from these groups must be tempered by the knowledge that they were not properly selected subsamples of the entire sample; hence there may be other causes operating beyond that on which we focus our attention.

a) In 65 cases the appraisers exceeded the minimum effort asked of them and went into the homes. We expected that their estimates would be more accurate, and that the r.m.s.(d) would be smaller. However, the r.m.s.(d) for these 65 cases turned out to be \$2,700 compared with \$3,100 for the entire sample. The appraisers' errors were not clearly increased by remaining outside the house.

b) In homes purchased during the calendar year prior to the interview, the respondent was asked what he actually paid for his home. We expect that the reports of the respondents were fairly close approximations to the true value at the time of purchase. The r.m.s.(d)

of \$1,900 is reliably smaller than the \$3,100 for the sample as a whole. One should not infer, however, that the entire \$1,900 is the result of errors by appraisers. For one thing, real estate values change with time, and up to a year might elapse between the purchase and the original interview, with several months more passing before the visit of the appraiser.

c) In the Surveys of Consumer Finances interviewers are instructed to make efforts to interview the head of the household rather than some other member. In the 91 cases where the interviews were taken from some other member of the household the r.m.s.(d) was not—contrary to expectations—larger than for the sample as a whole. In fact it was \$2,500 as against \$3,100 for the entire sample.

d) There were 59 cases where the head of the house was a female. For these the r.m.s.(d) was \$3,900, which appears to be reliably higher than for the entire sample.

The only important improvement in accuracy, then, was for respondents who purchased in the year prior to the survey. These respondents, as noted earlier, were asked what the property actually cost rather than their estimate of what it might be worth, hence, it is not surprising that their responses are close to the appraisers' estimates.

APPENDIX

The Model. The symbol r_i denotes the value recorded at the i^{th} home as a response in the interview survey; and a_i denotes the value assigned by the appraiser to the same home. The "true" (but unknown) value is y_i. Where there is little room for misunderstanding we shall drop the subscript i, and refer simply to r, a and y. The means over the entire population for the three sets of values may be designated by:

$$\overline{R} = E(r), \qquad \overline{A} = E(a), \qquad \overline{Y} = E(y). \tag{1}$$

The operator "E" denotes the "expected value of."[8] The variances of the three variables may be designated by:

$$V(r) = E(r - \overline{R})^2, \quad V(a) = E(a - \overline{A})^2, \quad V(y) = E(y - \overline{Y})^2. \tag{2}$$

[8] The means of the measurements r_i and a_i over a finite population would be variables also due to the errors of measurements. But we may treat $\overline{R}$ and $\overline{A}$ as constants if we consider them as resulting from a large number of reported measurements, or as coming from a large population. By confining ourselves to large populations we may also disregard any "finite population corrections" in our variance formulas. The terms used here are generally in accord with those in: M. H. Hansen, W. N. Hurwitz, and W. G. Madow, *Sampling Survey Methods and Theory* (New York: Wiley, 1953) II, Chap. 12.

Another good treatment of the topic of errors of response may be found in W. G. Cochran, *Sampling Techniques*, New York: Wiley, 1952, Chap. 13.

However, none of the sources known to us develop the model we need in terms of the differences $(r-a)$ of two sets of measurements, both subject to error. To what extent these non-sampling errors may be considered to be random variables is a complex problem which we shall have to leave untreated.

The quantity $(r_i - y_i)$ denotes the individual error of the response in the interview survey for the i^{th} home; and $(a_i - y_i)$ denotes the error in the appraiser's estimate for the same home. The difference between the two errors is equal to the difference of the two measurements:

$$d_i = (r_i - y_i) - (a_i - y_i) = (r_i - a_i). \tag{3}$$

Furthermore, let us call the mean value $(\bar{R} - \bar{Y})$ the response bias; $(\bar{A} - \bar{Y})$ the appraiser's bias; and the difference between the two biases is

$$\bar{D} = (\bar{R} - \bar{A}) = (\bar{R} - \bar{Y}) - (\bar{A} - \bar{Y}). \tag{4}$$

An important term in our model is the mean-square difference of the measurements:

$$\text{M.S.}(d) = E(d^2) = E(r - a)^2. \tag{5}$$

We also need the expression for the covariance between the differences in measurements and the appraiser's values:

$$\begin{aligned} \text{Cov}(da) &= E(d - \bar{D})(a - \bar{A}) = E(r - a - \bar{R} + \bar{A})(a - A) \\ &= E(r - a)(a - A), \end{aligned} \tag{6}$$

also:

$$\begin{aligned} \text{Cov}(da) &= E(r - \bar{R} - a + \bar{A})(a - \bar{A}) \\ &= E(r - \bar{R})(a - \bar{A}) - E(a - \bar{A})^2 = \text{Cov}(ra) - V(a). \end{aligned} \tag{6a}$$

With the above definitions, we may express the basic equation for our empirical investigations:

$$V(r) + \bar{D}^2 = V(a) + \text{M.S.}(d) + 2\text{Cov}(da). \tag{7}$$

For proof express $E(r - \bar{A})^2$ in two different ways:

$$E(r - \bar{A})^2 = E(r - \bar{R} + \bar{R} - \bar{A})^2 = V(r) + \bar{D}^2$$

and

$$(Er - \bar{A})^2 = E[(r - a) + (a - \bar{A})]^2 = \text{M.S.}(d) + V(a) + 2\text{Cov}(da).$$

Our model would be simpler if the appraiser gave the "true" value for every home, so that $a_i = y_i$; and the error equation would become

$$V(r) + (\bar{R} - \bar{Y})^2 = V(y) + E(r - y)^2 + 2\text{Cov}(r - y)(y).$$

Here $V(y)$ is the "true" sampling variance, i.e., the variance among the y_i, which are the "true" values of the homes; and

$$V(r) + (\bar{R} - \bar{Y})^2 - V(y) = E(r - y)^2 + 2\text{Cov}(r - y)(y).$$

is the increase in the total mean-square error due to errors of measurement. Similarly, the increase in the total mean-square error, due to the lesser accuracy of the r_i than the a_i, may be measured as

$$V(r) + \bar{D}^2 - V(a) = E(r - a)^2 - 2\mathrm{Cov}(r - a\,(a).$$

It is also interesting to note the relationship

$$E(r - a)^2 = E(r - y)^2 + E(a - y)^2 - 2E(r - y)(a - y).$$

The covariance $B(r-y)(a-y)$ of the two measurements may be positive or zero, but it is not likely to be an important negative quantity in the present instance. Therefore, the term $E(r-a)^2$ available in this study is likely to be larger than the mean-square error of response $E(r-y)^2$, by a quantity no greater than (but perhaps almost equal to) the mean-square error $E(a-y)^2$ of the appraiser's measurements.[9]

Although the results were obtained from a complex multi-stage sample, the discussion is given in terms of the composition of the response error for the individual homes which are the ultimate elements comprising the population. The expressions of the relative effects of the bias and of the variable error are given in terms of simple random samples. It is hoped that in this form the data will be of greater general interest and usefulness in planning other surveys. The calculations are based on the "naive" estimates from the pooled sample values; greater refinements did not seem to be warranted by the available data.[10]

The basic relationship shown in (7) may be expressed in terms of sample estimates as[11]

$$v(r) + \bar{d}^2 = v(a) + \mathrm{m.s.}(d) + 2\mathrm{cov}(da). \tag{8}$$

We have the following unbiased estimates:

$$\bar{r} = \frac{1}{n}\sum^{n} r, \qquad \bar{a} = \frac{1}{n}\sum^{n} a, \tag{9}$$

$$v(r) = \frac{1}{n-1}\sum^{n}(r - \bar{r})^2, \qquad v(a) = \frac{1}{n-1}\sum^{n}(a - \bar{a})^2, \tag{10}$$

$$\mathrm{cov}(ra) = \frac{1}{n-1}\sum^{n}(r - \bar{r})(a - \bar{a}), \tag{11}$$

---

[9] For the benefit of future researchers we should like to point out that an estimate of $E(a-y)^2$ could have been obtained had we assigned some of the homes to two appraisers each; we thought of this too late to carry out the necessary field work.

[10] For the sake of simplicity and because we have no measure of it, we disregard the correlation among the errors of individual homes, such as may be caused by interviewer bias.

[11] Because the responses were "weighted" to correct for the use of different sampling rates, the actual sample calculations were somewhat different from those shown here. For example, $\bar{r}$ was calculated as $\sum^{n} w_j r_j / \sum^{n} w_j$, where r_j is the response, and w_j is the assigned weight of the jth home in the sample. The calculation of the variances may be illustrated by

$$V(r) = \frac{n}{n-1}\,\frac{1}{\sum^{n} w_j}\,\sum^{n} w_j(r_j - \bar{r})^2.$$

$$\operatorname{cov}(da) = \operatorname{cov}(ra) - v(a) \tag{12}$$

$$\begin{aligned} \text{m.s.}(d) &= \frac{1}{n}\sum^{n} d^2 = \frac{1}{n}\sum^{n}(r-a)^2 \\ &= \frac{1}{n}\left[\sum^{n} r^2 + \sum^{n} a^2 - 2\sum^{n} ra\right]. \end{aligned} \tag{13}$$

Although $(\bar{r}-\bar{a})$ is an unbiased estimate of $\overline{D}$, $(\bar{r}-\bar{a})^2$ is not an unbiased estimate of $\overline{D}^2$; but $\bar{d}^2$ is an unbiased estimate where

$$\bar{d}^2 = (\bar{r}-\bar{a})^2 - \frac{1}{n}\left[v(r) + v(a) - 2\operatorname{cov}(ra)\right]. \tag{14}$$

This happens because

$$\begin{aligned} E(\bar{r}-\bar{a})^2 &= E\left[(\bar{r}-\overline{R}) - (\bar{a}-\overline{A}) + (\overline{R}-\overline{A})\right]^2 \\ &= V(\bar{r}) + V(\bar{a}) - 2\operatorname{cov}(\bar{r}\bar{a}) + \overline{D}^2. \end{aligned}$$

The unbiased estimate $\bar{d}^2$ has some advantages: together with the other unbiased estimates from the sample, it yields values for our error equation (8) which balance out exactly. However, it also has some disadvantages: it is a residual of sample values and it turns out to be negative sometimes—an embarrassing situation for the square of a real quantity. (One may decide to truncate the distribution of $\bar{d}^2$ at zero by substituting the value zero for all negative sample estimates. Alternately, one may use simply $(\bar{r}-\bar{a})^2$ with the knowledge that it has a positive bias of known magnitude.)

SOME CALCULATIONS ON THE DOLLAR VALUE OF THE HOUSE

The five terms of the basic equation (8) of the estimates of error components will be presented in this section for several situations. They will be given in units of \$1,000; since in these variance components the units are squared, a factor of 10^6 is needed to convert them to plain dollar values.

1) Our principal interest is in the components of the equation dealing with all the 568 cases:

$$32.65 + .10 = 26.69 + 9.58 - 3.52. \tag{15}$$

Note the relatively large m.s.(d) term which yields the $\sqrt{9.58 \times 10^6} = \$3{,}100$ estimate for the r.m.s.(d) between the two measurements on individual homes. But most of the discrepancies cancel leaving a much smaller net average error; the unbiased estimate of this bias is $\sqrt{.10 \times 10^6} = \$320$.

The total root-mean-square error of the responses is $\sqrt{v(r)} = \sqrt{32.65 \times 10^6} = \$5,700$ which is not much over the $\sqrt{v(a)} = \sqrt{26.69 \times 10^6} = \$5,200$ we would get from appraisers' estimates. Therefore, the practical surveyor may well be satisfied with the precision of the interview response—if the bias term is not too large.

The difference between the two variance terms is reduced by the sizeable negative covariance term (-3.52×10^6) between the difference in measurements $(r-a)$ and the appraisers' values. This is probably due in part to overestimates among the appraisers' high values and underestimates among their low values. The negative covariance is in accord with the gentle negative slope of curve A on Chart I.

2) If we allow the 13 gross coding errors (mentioned earlier) to stand uncorrected, the components are estimated as

$$37.35 + .04 = 26.69 + 15.81 - 5.11.$$

Thus over a third of the original m.s.(d) term of 15.81 was due to the 13 gross errors. However, the rise in the variance of the responses is more moderate (32 to 37). Moreover the estimate of the sample mean may be no worse off for these errors (ironically) because the bias term seems to be somewhat reduced. It seems that all these gross errors were in the direction of lowering the home-owners' estimates and, as noted above, home-owners suffer from a tendency to overestimate the value of their homes.

The effect of these coding errors on curve (A) in Chart I is to make the $(\bar{r}-\bar{a})$ values for the classes above \$10,000 more depressed and more irregular. Curve (B) of the r.m.s.(d) values is also disturbed above \$10,000: the curve becomes more irregular and the slope becomes greater (it seems to fit the line of $\sqrt{E(r-a)^2} = a/3$).

3) For the 65 cases where the appraiser went into the home the components are

$$37.07 - .09 = 34.31 + 7.29 - 4.62.$$

4) For the 61 cases where the response was in terms of the amount paid for a recently purchased home we have

$$23.14 - .05 = 21.76 + 3.67 - 2.34.$$

If we assume that the respondents gave the "true value" of their homes in these cases then we may accept this m.s.(d) term of 3.67 as a rough estimate of the appraisers' contribution to the discrepancy term.

5) For the 91 cases where the respondent was not the head of the household the values are

$$37.04 + .28 = 28.77 + 6.29 + 2.26.$$

6) For the 59 cases where the head of the household was a female the equation is

$$44.78 + .33 = 28.10 + 15.17 + 1.84.$$

RESULTS ON PROPORTIONS

When we deal with the proportion of cases which fall into any class interval our variables are binomial. The values of r_i and a_i are restricted to 0 and 1; and the value of $d_i = (r_i - a_i)$ is either 0, $+1$ or -1. The basic equation (8) of the estimated error components becomes:

$$\left[\frac{n}{n-1}\, p_r q_r\right] + \left[(p_r - p_a)^2 - \frac{1}{n-1}\,(p_r q_r + p_a q_a - 2p_{ra} + 2p_r p_a)\right] =$$
$$\left[\frac{n}{n-1} p_a q_a\right] + \left[p_r + p_a - 2p_{ra}\right] + \left[\frac{2n}{n-1}(p_{ra} - p_r p_a - p_a q_a)\right]. \quad (16)$$

Here p_r is the proportion of the homes placed into a specific frequency group by the responses to the interviews, while p_a is the proportion placed into that group by the appraisers' estimates. Also p_{ra} is the proportion placed into the same group both by respondent and by appraiser. Furthermore, $q_r = 1 - p_r$ and $q_a = 1 - p_a$. The equation for the 5,000–7,499 group would be, as read from the values of Table II:

$$\left[\frac{637}{636}\,(.196)(.804)\right] + \left[(.196 - .193)^2 - \frac{1}{636}[(.196)(.804)\right.$$
$$\left. + (.193)(.807) - 2(.083) + 2(.196)(.193)]\right]$$
$$= \left[\frac{637}{636}\,(.193)(.807)\right] + \left[.196 + .193 - 2(.083)\right]$$
$$+ \left[2\,\frac{637}{636}\,\{(.083) - (.196)(.193) - (.193)(.807)\}\right].$$

In Table IV, columns (1) to (5), we present the estimates of the five components of equation (16) for each of the classes shown in Table I. In column (6) we show the difference $(p_r - p_a)$ between the proportion assigned to each bracket in the surveys of respondents and appraisers. In column (7) we show the standard error of each difference shown in column (6).

TABLE IV

VALUES OF THE TERMS OF THE ERROR EQUATION (16) FOR THE PROPORTION IN EACH OF THE FREQUENCY CLASSES AS SHOWN IN TABLES I AND II

Frequency Group	Values of the Components of the Error Equation									(6) Difference Between Proportions Found (p_r-p_a)	(7) Standard Error of the Difference (p_r-p_a)
	(1) $v(r)$	+	(2) $\bar{d}^2$	=	(3) $v(a)$	+	(4) m.s.(d)	+	(5) 2cov(da)		
\$0– 2,499	.0282	−	.0000	=	.0225	+	.0320	−	.0263	+0.6%	0.7%
\$2,500– 4,999	.1140	−	.0002	=	.1184	+	.1240	−	.1286	−0.6%	1.4%
\$5,000– 7,499	.1578	−	.0003	=	.1560	+	.2230	−	.2215	+0.3%	1.9%
\$7,500– 9,999	.1690	+	.0004	=	.1842	+	.2360	−	.2508	−2.8%	1.9%
\$10,000–12,499	.1548	+	.0002	=	.1400	+	.2030	−	.1880	+2.3%	1.8%
\$12,500–14,999	.0609	+	.0004	=	.0804	+	.0930	−	.1121	−2.3%	1.2%
\$15,000–19,999	.0669	−	.0000	=	.0591	+	.0710	−	.0632	+0.9%	1.1%
\$20,000–29,999	.0273	−	.0000	=	.0216	+	.0340	−	.0283	+0.6%	0.7%
\$30,000 and over	.0148	−	.0000	=	.0138	+	.0130	−	.0120	+0.1%	0.4%
Not ascertained	.0530	−	.0001	=	.0449	+	.0930	−	.0850	+0.9%	1.2%
Two Illustrative Cumulated Groups:											
\$0– 7,500	.2296	−	.0003	=	.2287	+	.2090	−	.2084	+0.3%	1.8%
\$0–10,000	.2454	+	.0003	=	.2412	+	.2130	−	.2086	−2.5%	1.8%

Note that the m.s.(d) terms, denoting the variability due to the difference of the two responses, in column 4 of Table IV are large; generally they are as large as, or larger than, the $v(r)$ and $v(a)$ terms which ordinarily stand for sampling variability—shown in columns 1 and 3. One may be tempted to assume that this variability would be much less if larger groups were investigated; however, the two larger groups shown on the bottom two lines of Table IV, comprising respectively about 35 per cent and 60 per cent of the population, also have m.s.(d) terms almost as large as the $v(r)$ and $v(a)$ terms.

In spite of the large m.s.(d) the value of $[v(r)+\bar{d}^2]$ is hardly any larger than $v(a)$. This is due to the large negative covariance term. That is: there exists a large gross response variation but its net effect on variability is very small.

The net effect in terms of bias is even smaller. There is no bias term in column 2 which is reliable in terms of the standard error. If we average the ratios of the $\bar{d}^2$ values to the respective $v(r)$ values over the 10 classes we obtain .0005. In the calculations on the dollar mean the ratio of $\bar{d}^2$ term to the $v(r)$ term was .0030. Thus we may say that the bias term for the proportions remains undetected; and if it exists its effect on its total error is probably less than in the case of the dollar mean.

A COMPARISON OF STRATIFIED TWO-STAGE SAMPLING SYSTEMS

A. R. Sen, *Uttar Pradesh, India*

R. L. Anderson and A. L. Finkner, *North Carolina State College*

This paper deals with an empirical investigation of various stratified two-stage sampling systems for estimating totals of certain agricultural items of North Carolina. The 1940 Agricultural Census data were used for stratification, selection and estimation purposes. The observed data were the results of the 1945 Agricultural Census. Theory for the selection of n primary sampling units from a stratum with probability proportional to some measure of size but without replacement has already been developed by the senior author [11]. The principal contribution in this paper is the application of this theory to the selection of two primary sampling units without replacement from a stratum, where one of the units is selected with probability proportional to size and the other with equal probability. These results are compared with sampling systems (*i*) where both units are selected with probability proportional to size but with replacement and (*ii*) where an equal number of primary sampling units are selected but only one from each stratum.

1. INTRODUCTION*

The theory for selecting a single primary sampling unit (p.s.u.) per stratum with probability proportional to size (p.p.s.) in two-stage designs was developed by Hansen and Hurwitz [1] in 1943 and was applied to human populations. The theory for selecting more than one p.s.u. from each stratum with p.p.s. but with replacement was developed by Hansen and Hurwitz [2]. Theory for selecting two p.s.u.'s without replacement has been developed independently by Midzuno [8], Horvitz and Thompson [3], Narain [9], and Sen [10]. Both Midzuno and Sen generalized the Hansen and Hurwitz approach to sampling a combination of n elements of the universe with probability proportional to some measure of size of the combination. Sen [11] further derived an expression for an unbiased estimate of the variance of the estimate. The theory thus developed has been applied to four items of the North Carolina (N.C.) agricultural population. Results of the investigation will be presented in this paper.

One of the important results derived by Hansen and Hurwitz [1] was that selection of a p.s.u. from a stratum with p.p.s. was more efficient than selection with equal probability for a large class of populations. Their results were based on the between p.s.u. components only, the

within p.s.u. component being relatively small in all instances. Using a county as a p.s.u., Jebe [4] showed that the within p.s.u. component was relatively large for many agricultural populations of N.C., considering any reasonably practicable total sample size. He recommended the need for investigation of the township[1] as a p.s.u. This aspect of the sampling problem is also examined in this paper.

The values of four characteristics of the N.C. agricultural population have been studied. These are:

1. Number of non-white operators
2. Value of land and buildings
3. Number of days worked off farms
4. Total number of farms.

The sources of data were:

(*a*) U. S. Census of Agriculture 1945, vol. 1 part 16 (North Carolina and South Carolina),
(*b*) The 1945 Census of Agriculture for each minor civil division and the sample Census of Agriculture as available in I.B.M. punch cards.

For a general description of the population studied reference may be made to [7]. Most of the notations and terms employed in this paper have been used by Jebe [4]. For others, reference may be made to [8].

The principal objectives of this investigation were:

(*i*) to examine some applications of theory already developed for the selection of one p.s.u. per stratum,
(*ii*) to develop new theory for the selection of two p.s.u.'s per stratum,
(*iii*) to compare these two selection procedures empirically.

Some theoretical comparisons of the two selection procedures were made; however, no useful rules were found for indicating a preference for either procedure.

2. SAMPLING SYSTEMS WITH ONE P.S.U. SELECTED FROM EACH STRATUM

Three intensities of stratification were employed in this study, viz., 197 strata, [6], 98 strata, and 40 strata. The stratification was based on data provided by the 1940 Census of Agriculture. The counties of Dare and Swain were omitted from the study as they had only a small number of farms. The state was first divided into 197 strata following m.c.d.

[1] In North Carolina the township is also referred to as a minor civil division (m.c.d.).

lines and then 98 by combining two adjacent old strata, except that two of the new strata each contained about two and a half of the old strata. In this division care was taken to construct, as nearly as practicable, equal sized strata measured in terms of 1940 number of farms. Geographic contiguity of the m.c.d.'s within a stratum was maintained. The 40 strata were formed by combining contiguous counties of N.C. These forty strata were used for sampling designs with the county and with the m.c.d. as the p.s.u. However, equality of size of the strata was not feasible in this case.

The basic sampling design employed consists of two-stages of sampling in which

(a) one p.s.u. (i.e., a county or an m.c.d.) is selected from each stratum with equal probability or probability proportional to size, and
(b) a constant number of sub-sampling units (s.s.u.) (except in the 40 strata design) is selected at random from the s.s.u.'s located in the open country area of the p.s.u. selected in (a).

The s.s.u.'s are area segments delineated for the Master Sample of Agriculture project [5]. In the 40 strata design the total number of s.s.u.'s specified for the state was allocated proportionally among strata, i.e., proportional to the total number of s.s.u.'s in the open country portion of each stratum in 1945.

A summary of the various designs examined is given in Table 1. These designs are classified into five sampling systems A, B, C, D and E. A sampling system consists of the sample design and the method of estimation. A notation for designating the sampling systems discussed in this paper has been adopted. For simplicity this notation is confined to a single stratum, as is the discussion to follow. If Ω_p is the function designating the selection probabilities to be used and Y' is an estimator for the population total Y, for the characteristic of interest, then (Ω_p, Y') denotes the sampling system. For the two-stage sampling designs under consideration, where simple random sampling is always used in the second stage, Ω_p has been confined to the probabilities used for selecting the primary sampling units. To illustrate, if a single p.s.u. (say the ith) is selected and subsampled, let Y_i' be an unbiased estimate of Y_i its population total. Further suppose W_i is a weight function associated with Y_i such that W_iY_i' is an estimator for Y, the stratum total. If the p.s.u. is selected with probability proportional to X_i, then the sampling system is $[X_i/X, W_iY_i']$. The sampling systems A, B, C, D, and E are as follows:

A. (*i*) *p.s.u. selection:* p.p.s. with the number of farms in 1940 (F_{0i}) as the measure of size.
(*ii*) *s.s.u. selection:* designs 3, 4, 9, 10, 15, and 16 in Table 1.
(*iii*) *estimation:* ratio to the value of the characteristic in 1940 (X_i) for the p.s.u. sampled.
(*iv*) *sampling system* (*biased*):

$$\left[\frac{F_{0i}}{F_0},\ \frac{Y_i'}{X_i}X\right].$$

B. (*i*) *p.s.u. selection:* equal probability for each of the N p.s.u.'s in the stratum.
(*ii*) *s.s.u. selection:* designs 5, 6, 11, 12, 17, and 18 in Table 1.
(*iii*) *estimation:* same as in *A*.
(*iv*) *sampling system* (*biased*):

$$\left[\frac{1}{N},\ \frac{Y_i'}{X_i}X\right].$$

C. (*i*) *p.s.u. selection:* as in *A*.
(*ii*) *s.s.u. selection:* designs 3, 4, 9, 10, 15, and 16 in Table 1.
(*iii*) *estimation:* ratio to the number of farms in 1940 for the p.s.u. sampled.
(*iv*) *sampling system:*

$$\left[\frac{F_{0i}}{F_0},\ \frac{Y_i'}{F_{0i}}F_0\right].$$

D. (*i*) *p.s.u. selection:* p.p.s. with the value of the characteristic in 1940 as a measure of size.
(*ii*) *s.s.u. selection:* designs 1, 2, 7, 8, 13, and 14 in Table 1.
(*iii*) *estimation:* ratio to the value of the characteristic in 1940 for the p.s.u. sampled.
(*iv*) *sampling system:*

$$\left[\frac{X_i}{X},\ \frac{Y_i'}{X_i}X\right].$$

E. (*i*) *p.s.u. selection:* equal probability.
(*ii*) *s.s.u. selection:* designs 5, 6, 11, 12, 17, and 18 in Table 1.
(*iii*) *estimation:* estimated p.s.u. total weighted by the number of p.s.u.'s in the stratum.
(*iv*) *sampling system:*

$$\left[\frac{1}{N},\ NY_i'\right].$$

It should be clear that the method of selecting the p.s.u. remains the same for each of the four characteristics observed when sampling systems *A*, *B*, *C*, or *E* are used. This is not the case for system *D*, in which the probability of selection depends on the value of the characteristic in 1940 for the p.s.u. sampled. Systems *A*, *B*, *C*, and *E*, therefore, could be recommended for a general purpose survey, i.e. where the purpose of the survey is to estimate the totals of several characteristics. System *D*, however, could be recommended only where information is de-

sired on only one characteristic or where additional characteristics observed must be subordinated in favor of a single characteristic.

Expressions for mean square errors, i.e. the between and within components of variance and the bias terms, for the various sampling sys-

TABLE 1

SAMPLING DESIGNS FOR ONE P.S.U. PER STRATUM

Design No.	Method of selection of p.s.u.	No. of s.s.u.'s per selected p.s.u.	Sampling Rate %
	197 Strata		
1	P.P.S.-Value of characteristic in 1940	2	1
2	P.P.S.-Value of characteristic in 1940	4	2
3	P.P.S.-Number of farms in 1940	2	1
4	P.P.S.-Number of farms in 1940	4	2
5	Equal probability	2	1
6	Equal probability	4	2
	98 Strata		
7	P.P.S.-Value of characteristic in 1940	4	1
8	P.P.S.-Value of characteristic in 1940	8	2
9	P.P.S.-Number of farms in 1940	4	1
10	P.P.S.-Number of farms in 1940	8	2
11	Equal probability	4	1
12	Equal probability	8	2
	*40 Strata**		
13	P.P.S.-Value of characteristic in 1940	5	0.5
14	P.P.S.-Value of characteristic in 1940	20	2
15	P.P.S.-Number of farms in 1940	5	0.5
16	P.P.S.-Number of farms in 1940	20	2
17	Equal probability	5	0.5
18	Equal probability	20	2

* No. s.s.u.'s per p. s. u. is an average figure.

tems are given in the appendix. The general procedure for derivation will be indicated here. Consider the sampling system (Ω_p, W_iY_i') as illustrated above. It is easy to see that, in general, this system is biased for the estimation of $\bar{Y}$, because

$$E(W_i\cdot Y_i') = E_1\cdot E_2(W_iY_i') = E_1(W_iY_i),$$

where E_1 refers to expectation over the first stage of sampling and E_2 over the second stage. This estimate is unbiased if, and only if, $\Omega_p = 1/W_i$. In particular, let Ω_p be equal to P_i where $\sum_i P_i = 1$. Thus

$$E_1(W_iY_i) = \sum_i P_i\cdot W_i\cdot Y_i.$$

This sum equals Y if, and only if, $P_i = 1/W_i$, for all i. Hence, $[1/W_i, W_i \cdot Y_i']$ is an unbiased sampling system. A general expression for the mean square error for systems in the class considered is given by

$$E[W_iY_i' - Y]^2 = \underbrace{E[W_i^2(Y_i' - Y_i)^2]}_{\text{Within variance}} + \underbrace{E[W_iY_i - E(W_iY_i)]^2}_{\text{Between variance}} + \underbrace{[E(W_iY_i) - Y]^2}_{\text{Square of Bias}} \cdots (1)$$

The mean square error for a particular system is obtained by substituting the corresponding values of W_i for the system in equation (1) above and taking the expectations according to the probability function, Ω_p, e.g. for system A, $W_i = X/X_i$ and $\Omega_p = F_{0i}/F_0$. The mean square error for the state estimate is found by summing (1) over all strata, except that the bias is first summed over all strata and the total bias is squared.

3. ANALYSIS OF VARIANCE COMPONENTS FOR SYSTEMS A, B, C, D, AND E

In order to compare sampling systems A, B, C, D, and E the estimated $(\text{C.V.})^2 \times 10^4$, where $(\text{C.V.})^2$ is the estimated mean square error divided by Y^2, are presented in Tables 2 and 3. The within p.s.u. components are shown in Table 2, the total error in Table 3. The between components of error may be obtained by subtraction. This latter component includes both the between p.s.u. variance and the bias contribution for systems A and B. In calculating the between component contributions, exact expressions for the expected values have been obtained. Since information was available for only a one in eighteen systematic sample of the Master Sample segments within each county, considerable difficulty was experienced in obtaining estimates of the within p.s.u. component of the total error. Furthermore, this sample embraced incorporated, unincorporated and open country areas. In this connection Jessen [5, p. 536] says, "The areas into which the open country zone was partitioned serve as units for sampling either or both farms and persons whether farm or non-farm. This portion of the sample is as useful, therefore, for a sample census of population as for a sample census of farms. This dual purpose sampling unit is feasible only in the open country, where the majority of the families are engaged in farming." Hence only data for the open country area were used in the estimation of within p.s.u. variances. The Master Sample segment summary cards which belonged to incorporated or unincorporated places or in a few cases to open country areas falling within

metropolitan districts, or which formed subunits of multiple units had to be dropped. As the open country area included about 90 per cent of the total number of segments for North Carolina, the within p.s.u. variation for this section of the state is fairly representative of the entire state. Strictly speaking, however, the conclusions drawn are valid

TABLE 2

ESTIMATED $(\text{C.V.})^2 \times 10^4$ FOR THE WITHIN P.S.U. COMPONENT OF ERROR

SAMPLING SYSTEM*	A†		B†		C		D		E	
CHARACTERISTIC										
Sampling Rate (%)	1	2	1	2	1	2	1	2	1	2
197 Strata (m.c.d.)										
No. of Non-White Operators	118	55	135	63	44	21	55	26	47	22
Value of Land and Buildings	15	7	17	8	14	6	14	7	15	7
No. of Days Worked off Farms	129	60	154	72	16	8	23	11	18	8
Total No. of Farms	15	7	16	7	15	7	15	7	16	8
98 strata (m.c.d.)										
No. of Non-White Operators	131	57	151	65	42	19	55	25	47	21
Value of Land and Buildings	17	7	18	8	13	6	14	6	14	6
No. of Days Worked off Farms	166	71	193	83	15	7	24	11	17	8
Total No. of Farms	14	6	15	7	14	6	14	6	16	8
Sampling Rate (%)	0.5	2	0.5	2	0.5	2	0.5	2	0.5	2
40 Strata (m.c.d.)										
No. of Non-White Operators	296	36	345	38	76	11	104	15	85	14
Value of Land and Buildings	32	4	35	4	25	4	27	4	20	5
No. of Days Worked off Farms	681	83	835	99	30	4	54	8	33	5
Total No. of Farms	27	4	28	4	27	4	27	4	30	5
40 Strata (County)										
No. of Non-White Operators	145	35	152	35	97	23	106	25	115	27
Value of Land and Buildings	41	10	42	10	38	9	39	9	44	10
No. of Days Worked off Farms	67	16	71	17	41	10	48	11	47	11
Total No. of Farms	36	9	37	9	36	9	36	9	40	10

* See Section 2 for definitions of sampling systems.

† The bias contribution to the total error is included in the between p.s.u. $(\text{C.V.})^2$.

for the open country area only. It was further assumed that the estimate of the within county and m.c.d. variation obtained from systematic sampling is approximately equal to that of a random sample if an equal number of segments are selected from the same population.

The within p.s.u. component required the estimation of both within county and within m.c.d. variation. For a few of the counties and for a great many more of the m.c.d.'s, the number of s.s.u.'s available for estimating the within variation was too small to provide efficient estimates. Furthermore, of the 941 m.c.d.'s used in this investigation,

342 provided no estimates of the within variation, since data were available in each of these for either none or only one s.s.u.

The method of estimation used consisted of pooling the observed within p.s.u. variances for contiguous p.s.u.'s so that the resultant estimates were based on more degrees of freedom, thus increasing their

TABLE 3

ESTIMATED $(C.V.)^2 \times 10^4$ FOR THE TOTAL COMPONENT OF ERROR

SAMPLING SYSTEM*	A†		B†		C		D		E	
CHARACTERISTIC										
Sampling Rate (%)	1	2	1	2	1	2	1	2	1	2
197 Strata (m.c.d.)										
No. of Non-White Operators	137	74	168	96	52	29	61	32	65	40
Value of Land and Buildings	21	13	27	18	19	12	18	11	25	18
No. of Days Worked off Farms	681	612	750	668	63	55	68	56	76	66
Total No. of Farms	17	9	19	11	17	9	17	9	23	15
98 Strata (m.c.d.)										
No. of Non-White Operators	161	87	208	122	65	41	69	38	89	63
Value of Land and Buildings	28	19	34	24	27	20	23	15	41	33
No. of Days Worked off Farms	1168	1074	1405	1294	133	124	122	108	166	157
Total No. of Farms	19	11	22	13	19	11	19	11	34	26
Sampling Rate (%)	0.5	2	0.5	2	0.5	2	0.5	2	0.5	2
40 Strata (m.c.d.)										
No. of Non-White Operators	371	112	463	157	151	86	141	52	201	130
Value of Land and Buildings	62	35	70	39	68	46	53	29	113	89
No. of Days Worked off Farms	4601	4004	5235	4500	386	361	345	299	505	477
Total No. of Farms	40	17	49	24	40	17	40	17	86	61
40 Strata (County)										
No. of Non-White Operators	150	39	161	44	112	39	109	28	195	107
Value of Land and Buildings	42	11	44	12	45	16	40	11	93	60
No. of Days Worked off Farms	114	63	[illegible]	108	152	121	96	60	201	165
Total No. of Farms	38	10	44	16	38	10	38	10	82	53

* See Section 2 for definitions of sampling systems.

† The bias contribution to the total error is included in the between p.s.u. $(C.V.)^2$.

stability. This method assumes, of course, that the true within p.s.u. variances do not vary for those p.s.u.'s which were combined. Since this assumption is not generally valid, the estimates obtained of the within variation may be slightly biased, thus affecting any comparisons amongst two or more sampling systems. This point needs further investigation. Three sets of pooled estimates of within p.s.u. variances were worked out.

(a) The 40 strata with the county as the p.s.u. were grouped into 20

strata,[4] each new stratum consisting of two contiguous old strata. The within county variances for each new stratum were pooled to yield an over-all estimate of variance for the stratum. This estimate was used for each county in the stratum.

(b) For each of the 20 strata obtained in (a) the within m.c.d. variances were pooled to yield another estimate of the within p.s.u. variance for each stratum when the designs using 40 strata with the m.c.d. as the p.s.u. were studied.

(c) The 98 strata with the m.c.d. as the p.s.u. were pooled into 20 strata such that the stratification was almost the same as in (a) and (b). The within m.c.d. variances within each new stratum were pooled to yield an estimate of within p.s.u. variance for the stratum. This was applied to each m.c.d. within each stratum, when the designs using 98 and 197 strata were studied.

The within county contributions to total error are considerably greater than the between contributions for all the sampling systems. Even with the m.c.d. as the p.s.u., this contribution is a very important factor. As pointed out in the introduction, this fact was also observed by Jebe [4]. Hence it might be feasible to consider a delineation of s.s.u.'s which are more homogeneous than the present Master Sample segments.

It can be seen from Table 3 that the total $(C.V.)^2$ for all the characteristics and for all the sampling systems using 40 strata is less where the p.s.u. is a county than when the p.s.u. is an m.c.d. This difference is marked for number of days worked off farms for all the sampling systems, particularly A and B. One reason for this marked difference is the smaller within contribution when the p.s.u. is a county. This seemingly anomalous result arises from the instability of the weights $(W_i = X/X_i)$ for number of days worked off farm (and also for number of non-white operators) when the p.s.u. is an m.c.d. In many m.c.d.'s, X_i is very small even though the probability of selection (F_{0i}/F_0) is large; since the within component is $E[W_i^2(Y_i' - Y)^2]$, a very small value of X_i (large value of W_i) can have a tremendous effect on the within contribution. The W_i are much more stable when the p.s.u. is a county.

Six main comparisons of sampling systems have been made on the basis of the relative errors shown in Tables 2 and 3. These comparisons can be divided into two groups.

---

[4] These will be described below in the discussion on the selection of two or more p.s.u.'s from a stratum.

(a) Comparisons of sampling systems differing in the method of selection

It would appear from the tables that the biased system A, where selection is p.p.s. to number of farms in 1940 but the estimator is the ratio to the value of the characteristic in 1940, is more efficient for all the characteristics and for all stratifications than the biased system *B*, where selection is made with equal probability with the same estimator. The gain in efficiency is, however, very small when 40 counties are used for estimating the number of non-white operators and value of land and buildings. This does not mean that selection with equal probability is almost as effective as p.p.s. selection, since the appropriate comparison is on only the between p.s.u. components of variance. For the between p.s.u. components of variance, system *A* showed considerable gain in efficiency relative to system *B*. System *D*, where selection is made with p.p.s. to value of the characteristic in 1940, is more efficient than system *A* or *B* for all characteristic totals, except for total number of farms where systems *A* and *D* are identical. The relative efficiency of system *D* is highly pronounced for estimating the number of non-white operators and number of days worked off farms when the m.c.d. is used as the p.s.u.

(b) Comparisons of sampling systems differing in the method of estimation

The unbiased system *C*, in which selection is p.p.s. to number of farms in 1940 and the estimator is ratio to number of farms in 1940, is generally more efficient than the biased system *A*, when the m.c.d. is used as the p.s.u. The gain in efficiency is most pronounced for number of days worked off farms and is identically unity for total number of farms. With the county as the p.s.u., the relative efficiency of *C* to *A* is considerably reduced and is in fact less than unity for value of land and buildings and number of days worked off farms.

The unbiased system *E*, where selection is with equal probability and estimation is accomplished by a simple expansion of the estimated p.s.u. total by the number of p.s.u.'s in the stratum, is generally more efficient than the biased system *B* for estimating the number of non-white operators and the number of days worked off farms using the m.c.d. as the p.s.u. However, the situation is reversed for value of land and buildings and total number of farms, for which the correlations between the 1940 and 1945 values are both high. When the county is used as the p.s.u., system *B* is more efficient than system *E* for estimating totals of all the characteristics.

As regards the between components of variance, the county is always a better p.s.u. for all the characteristics for biased systems *A* and *B* compared to unbiased systems *C* and *E*. The reduction in the sampling

error portion of this component for the biased systems A and B, due to the high correlations between the 1940 and 1945 values of each of the characteristics with the county as a p.s.u., more than compensated for the loss due to the bias contribution.

Because of the high C.V. for number of days worked off farms, it is very difficult to find a sampling system which will give acceptable results for all four characteristics.

Any one characteristic total can be estimated satisfactorily by use of system D for that characteristic; however, this procedure will not give satisfactory results for the other three characteristics. Suppose the standard is set up that, using a two per cent sample, each characteristic total shall be estimated with no more than a ten per cent C.V., i.e. an accuracy of estimation to within 20 per cent of the item total with 95 per cent confidence. None of the sampling systems will provide this accuracy for all types of stratification and p.s.u. considered here. However if 40 strata with the county as the p.s.u. are used, system A will meet the standard and system B will almost meet it. If 197 strata are used, with a two per cent sampling rate, systems C and E will provide estimates of the characteristic totals within an eight per cent C.V.

None of the systems considered will provide an estimate of the number of days worked off farms within a five per cent C.V. Hence, it was deemed advisable to investigate the possibility of sampling from each of the 98 counties (this would be single stage sampling). If all 98 counties were sampled, the between component of the total error would vanish; hence, the total error would be simply the within component. This within component, using the 98 counties as strata, was determined easily from the calculations of the within component for system E for the 40 strata with the county as the p.s.u. In order to use the existing calculations, it is noted that, except for weighting factors, the within component of a $\frac{1}{2}$ per cent sample using 40 counties corresponds to the total error for about a 1.25 per cent sample using all 98 counties (actually about 5 s.s.u.'s per county) and a 2 per cent sample of 40 counties corresponds to a 5 per cent sample of all 98 counties. The estimated per cent C.V. using the 98 counties as strata are presented below.

Characteristic	Sampling Rate	
	1.25%	5%
Number of Non-White Operators	7.2	3.5
Value of Land and Buildings	4.3	2.1
Number of Days Worked off Farms	4.5	2.2
Total Number of Farms	4.3	2.1

From these results it appears that a sampling rate of 2.5 per cent would be sufficient to estimate each characteristic within a five per cent C.V. if all 98 counties were included in the sample.

4. SELECTION OF TWO PRIMARY SAMPLING UNITS FROM A STRATUM

Only one intensity of stratification was employed in this study. The state was divided into 20 strata. Each stratum was formed by combining two contiguous strata of the 40 strata with the county as the p.s.u. discussed under selection of one p.s.u. from a stratum. Four sampling systems F, G, H, and K were considered for each of the four characteristics. These are:

F. (*i*) *p.s.u. selection:* the first with p.p.s. to the value of the characteristic in 1940 and the other with equal probability but without replacement from the remaining p.s.u.'s in the stratum.

(*ii*) *s.s.u. selection:* random and independently from each of the p.s.u.'s selected in (*i*) above. The number of s.s.u.'s selected from each of the 20 strata was proportional to the total number of s.s.u.'s in the open country area of the stratum. Two subsampling rates were used, i.e. 0.5 and 2.0 per cent.

(*iii*) *estimation:* ratio of the estimated total of the characteristic to the total value of the characteristic in 1940 for the p.s.u.'s selected in (*i*).

(*iv*) *sampling system:*

$$\left[\frac{X_i + X_j}{(N-1)X},\quad \frac{Y_i' + Y_j'}{X_i + X_j} X\right].$$

G. (*i*) *p.p.s. selection:* the first with p.p.s. to the number of farms in 1940 and the other with equal probability but without replacement from the remaining p.s.u.'s in the stratum.

(*ii*) *s.s.u. selection:* same as in F (*ii*).

(*iii*) *estimation:* ratio of the estimated total of the characteristic to the total number of farms in 1940 for the p.s.u.'s selected in G (*i*).

(*iv*) *sampling system:*

$$\left[\frac{F_{0i} + F_{0j}}{(N-1)F_0},\quad \frac{Y_i' + Y_j'}{F_{0i} + F_{0j}} F_0\right].$$

H. (*i*) *p.s.u. selection:* the two p.s.u.'s each with p.p.s. to the value of the characteristic in 1940 but with replacement.

(*ii*) *s.s.u. selection:* same as in F (*ii*).

(*iii*) *estimation:* average of the ratios of the estimated totals of the characteristic to the corresponding value of the characteristic in 1940 for the p.s.u.'s selected.

(*iv*) *sampling system:*

$$\left[\frac{2X_iX_j}{X^2},\quad \left(\frac{Y_i'}{X_i} + \frac{Y_j'}{X_j}\right)\frac{X}{2}\right].$$

K. (*i*) *p.s.u. selection:* two p.s.u.'s each with p.p.s. to the total number of farms in 1940 but with replacement.

(*ii*) *s.s.u. selection:* same as in F (*ii*).

(*iii*) *estimation:* average of the ratios of the estimated total of the characteristic to the total number of farms in 1940 for the p.s.u.'s selected.

(*iv*) *sampling system:*

$$\left[\frac{2F_{0i}F_{0j}}{F_0^2},\ \left(\frac{Y_i'}{F_{0i}}+\frac{Y_j'}{F_{0j}}\right)\frac{F_0}{2}\right].$$

The method of selecting the p.s.u.'s for systems F and H depends on the value of the characteristic for these units in 1940. Therefore, these systems would be useful only for specific purpose surveys. On the other hand, systems G and K for which the method of selection of the p.s.u.'s is based on number of farms in 1940 would be suitable for general purpose surveys. Expressions of variances for each of the sampling systems described above are given in the Appendix; however, the procedure for arriving at an expression for the variance will be indicated for one of the systems. Consider the unbiased sampling system

$$F:\left[\frac{X_i+X_j}{(N-1)X},\quad \frac{Y_i'+Y_j'}{X_i+X_j}X\right].$$

$$\begin{aligned}
\text{Var.}\left[\frac{Y_i'+Y_j'}{X_i+X_j}X\right] &= E\left\{\left[\frac{Y_i'+Y_j'}{X_i+X_j}X-Y\right]^2\right\}\\
&= E\left\{\left[\frac{Y_i'+Y_j'}{X_i+X_j}X-\frac{Y_i+Y_j}{X_i+X_j}X\right.\right.\\
&\qquad\left.\left.+\frac{Y_i+Y_j}{X_i+X_j}X-Y\right]^2\right\}\\
&= \underbrace{\left\{\sum_{i<j}\sum\frac{1}{(N-1)}\frac{(Y_i+Y_j)^2}{(X_i+X_j)}X-Y^2\right\}}_{\text{Between}}\\
&\quad+\underbrace{\left\{\sum_{i<j}\sum\frac{(Z_i+Z_j)X}{(N-1)(X_i+X_j)}\right\}}_{\text{Within}},
\end{aligned}$$

where $Z_i=M_i(M_i-m_i)\sigma_i^2/m_i$.

5. ANALYSIS OF VARIANCE COMPONENTS FOR SYSTEMS F, G, H, AND K

This section will deal with the analysis of the variance components of the sampling systems described in Section 4. The results of the analysis for systems C and D, where only one p.s.u. is selected from a stratum, are also presented here to facilitate comparative study with systems in which two p.s.u.'s are selected from a stratum. In this study,

as before, the square of the coefficient of the variation, $(C.V.)^2$, will be used for the total error. Calculations for both the between and within components of error have been made only for systems *F* and *G*, in which the selection of p.s.u.'s is made without replacement. For systems *H* and *K*, in which the p.s.u.'s are selected with replacement, calculations have been made on the between p.s.u. components of the error only.

From Tables 4 and 5 it would appear that systems *F* and *G* are, respectively, more efficient than systems *D* and *C* in estimating number of days worked off farms and are highly so on the between components

TABLE 4

BETWEEN COMPONENTS OF $(C.V.)^2 \times 10^4$ FOR SAMPLING SYSTEMS C, D, F, G, H, AND K*

Sampling System	G	K	C	F	H	D
Measure of Size	1940 No. of Farms			1940 Value of Char.		
CHARACTERISTIC						
No. of Non-White Operators	16	42	16	3	8	3
Value of Land and Buildings	10	27	7	2	5	1
No. of Days Worked off Farms	82	252	111	43	117	48
Total No. of Farms	1	4	2	1	4	2

* For definition of sampling systems C and D, see Section 2 and for F, G, H, and K, see Section 4. $(C.V.)^2 \times 10^4$ rounded to the nearest integer but $(C.V.)^2$ calculated correct to the sixth decimal figure.

TABLE 5

TOTAL $(C.V.)^2 \times 10^4$ FOR SAMPLING SYSTEMS *C*, *D*, *F*, AND *G*

Measure of Size	1940 No. of Farms				1940 Value of Char.			
Sampling System	G		C		F		D	
Sampling Rate %	0.5	2	0.5	2	0.5	2	0.5	2
CHARACTERISTIC								
No. of Non-White Operators	125	42	112	39	115	30	109	28
Value of Land and Bldgs.	52	20	45	16	45	12	40	11
No. of Days Worked off Farms	127	93	152	121	92	55	96	60
Total No. of Farms	41	11	38	10.5	41	11	38	10.5

of variance. With an increase in the sub-sampling rate from 0.5 to 2.0 per cent, there is an appreciable reduction in the total error. Systems *F* and *G* have an additional advantage over systems *D* and *C*, respectively, in that they would provide estimates of the between components of error from the sample. On the between components of variance, sys-

tems F and G are, respectively, more efficient than systems D and C in estimating the total number of farms. This seems to be a paradox, since the between p.s.u. (C.V.)^2 for any of 20 strata for systems F and G is expected to include in addition some between strata variance (of the original 40 strata). The method of estimation introduced in estimating the population total by the selection of two p.s.u.'s seems to cut down the variance of the estimated total, because the ratio $(Y_1'+Y_j')/(X_i+X_j)$ may be less variable than Y_i'/X_i. This reduction is naturally more pronounced where the between p.s.u. contribution is high, i.e., with number of days worked off farms for which the reduction in variance due to the method of estimation more than compensates for the increase in variance due to the inclusion of some between strata variation.

The between components of the total error for systems F and G, where selection is made without replacement, are highly efficient compared to the same components for systems H and K where selection is made with replacement.

With a two per cent sample, it is estimated that system G would provide an estimate of each characteristic total within a ten per cent C.V. and within a seven per cent C.V. for all characteristics excepting number of days worked off farms. For a specific purpose survey, system F should be used.

6. CONCLUSIONS

Of the four characteristics investigated in this paper, one was influenced greatly by war-time activities between 1940 and 1945. None of the sampling systems studied was suitable to estimate this item, number of days worked off farms, with seven per cent accuracy when a sampling rate of two per cent is used. If a stratified sample using the 98 counties as strata is substituted for a two-stage sampling system, a 1.25 per cent sampling rate would be sufficient to estimate this item with a five per cent C.V. The conclusions which follow relate to the other three characteristics only, namely number of non-white operators, value of land and buildings and total number of farms.

For a general purpose survey to estimate characteristic totals, system C is recommended. This system will provide estimates for all the types of stratification considered with no more than a nine per cent C.V. if a two per cent sampling rate is used.

Among the four types of stratification and p.s.u. considered in this study, two of them, namely, 40 strata using the county or 197 strata using the m.c.d. as the p.s.u., were found to be most suitable from the point of view of total error and are recommended for estimating the

characteristic totals. With a two per cent sampling rate using system C and either of these types of stratification, estimates of the characteristic totals could be obtained within a C.V. of six per cent.

The within p.s.u. component of error is not reduced to the same extent as the between p.s.u. component is increased, when the m.c.d. is used as the p.s.u. instead of the county with 40 strata. Even with the m.c.d. as the p.s.u., the within contribution to the total error is a very important factor for any reasonably practicable sample size.

The efficiencies of p.p.s. and equal probability selection are compared in systems A and B. The method of estimation, ratio to the value of the characteristic in 1940, is the same in both systems. For all cases investigated, system A (p.p.s. selection of p.s.u.'s with number of farms in 1940 as the measure of size) was found to be superior to system B (selection of p.s.u.'s with equal probability).

Although system D, in which selection is p.p.s. to the 1940 value of the characteristic, is impracticable when it is desired to estimate the totals of a number of characteristics from a single sample, it is generally suitable for specific purpose surveys, where one is interested in only a single characteristic. However, system C was found to be nearly as efficient as system D in estimating totals for the characteristics studied in this paper. This may not be true in general.

The choice of a sampling system has been based on the assumption of equal costs per schedule. There is a real need for using a more realistic cost function, but this would necessitate acquiring adequate and accurate information on the various cost factors, e.g., cost of travel and cost of enumeration in survey designs.

The principal contribution in this paper is the examination of a sampling design for the selection from a stratum of two p.s.u.'s, the first with p.p.s. and the second with equal probability from the remaining p.s.u.'s. The advantage of this design over one involving selection of a single p.s.u. from each stratum is that it permits an unbiased estimate [11] of the sampling error from sample data. In some cases, in addition, the system involving the selection of two p.s.u.'s per stratum may be more efficient than systems permitting the selection of only one p.s.u. from each of twice as many strata. It appears that this will be true when there is extreme variability among the p.s.u.'s, such as in the case of number of days worked off farms. The between p.s.u. variance for the three remaining items studied is much lower, but even in these cases the efficiencies of the systems selecting one p.s.u. per stratum do not greatly exceed those in which two p.s.u.'s per stratum are selected without replacement.

A considerable part of the material in this paper is taken from the

senior author's thesis submitted in partial fulfilment of the requirements of the Ph.D. degree in Experimental Statistics, North Carolina State College, 1952. Correspondence with E. H. Jebe of Iowa State College and Morris Hansen of the Bureau of Census helped in the general orientation and understanding of certain aspects of the problem. The computational work involved in this study was enormous and would have remained incomplete but for the special operations devised by members of the I.B.M. staff. The authors wish to express their appreciation to D. G. Horvitz for offering many valuable suggestions in the final preparation of this paper.

REFERENCES

[1] Hansen, Morris H., and Hurwitz, William N., "On the Theory of Sampling from Finite Populations," *Annals of Mathematical Statistics*, 14 (1943), 332–62.

[2] Hansen, Morris H., and Hurwitz, William N., "On the Determination of Optimum Probabilities in Sampling," *Annals of Mathematical Statistics*, 20 (1949), 426–32.

[3] Horvitz, D. G., and Thompson, D. J., "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47 (1952), 663–85.

[4] Jebe, Emil H., "Estimation for Sub-sampling Designs Employing the County as a Primary Sampling Unit," *Journal of the American Statistical Association*, 47 (1952), 49–70.

[5] Jessen, Raymond J., "The Master Sample Project and Its Uses in Agricultural Economics," *Journal of Farm Economics*, 29 (1947), 531–40.

[6] Kastenbaum, Marvin A., "On Sampling for State Estimates of the Farm Population in North Carolina Using the Township as a Primary Sampling Unit," Institute of Statistics Mimeograph Series, No. 29. North Carolina State College, Raleigh, N. C. (1950).

[7] Madow, Lillian H., "On the Use of the County as the Primary Sampling Unit for State Estimates," *Journal of the American Statistical Association*, 45 (1950), 30–47.

[8] Midzuno, Hiroshi, "An Outline of the Theory of Sampling Systems," *Annals of the Institute of Statistical Mathematics* (Japan), (1950), 149–56.

[9] Narain, R. D., "On Sampling Without Replacement with Varying Probabilities," *Journal of the Indian Society of Agricultural Statistics*, 3 (1951), 169–74.

[10] Sen, Amode Ranjan, "Present Status of Probability Sampling and its Use in the Estimation of Farm Characteristics," Paper presented at the joint meeting of the Econometric Society with the Institute of Mathematical Statistics, Minneapolis, Minn., Sept. 1951. [Abstract in *Econometrica*, 20 (1952), 103.]

[11] Sen, Amode Ranjan, "Further Developments of the Theory and Application of the Selection of Primary Sampling Units, with Special Reference to the North Carolina Agricultural Population," Ph.D. Thesis, Library, North Carolina State College.

APPENDIX

The expressions given below are the mean square errors for the estimates for each of the nine sampling systems presented in this paper (Section 2 and Section 4). As before A, B, C, D, and E stand for the systems in which one p.s.u. is selected from a stratum and F, G, H, and K for the systems in which two p.s.u.'s are selected from a stratum. For simplicity of notation only the results for a single stratum are presented. These results when summed over all strata will provide the mean square error of the estimate for the entire state, noting that the bias is first summed over all strata and this sum is squared.

Sampling system A:

$$\left[\frac{F_{0i}}{F_0},\ \frac{Y_i'}{X_i}X\right]$$

$$V_A = X^2\left\{\sum_i \frac{F_{0i}}{F_0}\cdot\frac{1}{X_i^2}\cdot M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}\right\}$$

Within p.s.u. variance

$$+ X^2\left\{\sum_i \frac{F_{0i}}{F_0}\left[\frac{Y_i}{X_i}\right]^2 - \left[\sum_i \frac{F_{0i}}{F_0}\frac{Y_i}{X_i}\right]^2\right\}$$

Between p.s.u. variance

$$+ X^2\left\{\left[\sum_i \frac{F_{0i}}{F_0}\frac{Y_i}{X_i} - \frac{Y}{X}\right]^2\right\},$$

(Bias)^2

where σ_i^2 = variance for single s.s.u.'s selected at random within the ith p.s.u.

Sampling system B:

$$\left[\frac{1}{N},\ \frac{Y_i'}{X_i}X\right]$$

$$V_B = X^2\left\{\sum_i \frac{1}{N}\cdot\frac{1}{X_i^2}\cdot M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}\right\}$$

Within variance

$$+ X^2\left\{\sum_i \frac{1}{N}\left[\frac{Y_i}{X_i}\right]^2 - \left[\sum_i \frac{1}{N}\frac{Y_i}{X_i}\right]^2\right\}$$

Between variance

$$+ X^2\left\{\left[\sum_i \frac{1}{N}\frac{Y_i}{X_i} - \frac{Y}{X}\right]^2\right\},$$

(Bias)^2

where σ_i^2 is defined as for V_A.

Sampling system C:

$$\left[\frac{F_{0i}}{F_0},\quad \frac{Y_i'}{F_{0i}}F_0\right]$$

$$V_C = \underbrace{F_0\left\{\sum_i \frac{1}{F_{0i}}\cdot M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}\right\}}_{\text{Within variance}} + \underbrace{\left\{F_0\sum_i \frac{Y_i^2}{F_{0i}} - Y^2\right\}}_{\text{Between variance}},$$

where σ_i^2 is defined as for V_A.

Sampling system D:

$$\left[\frac{X_i}{X},\quad \frac{Y_i'}{X_i}X\right]$$

$$V_D = \underbrace{X\left\{\sum_i \frac{1}{X_i}\cdot M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}\right\}}_{\text{Within variance}} + \underbrace{\left\{X\sum_i \frac{Y_i^2}{X_i} - Y^2\right\}}_{\text{Between variance}},$$

where σ_i^2 is defined as for V_A.

Sampling system E:

$$\left[\frac{1}{N},\ NY_i'\right]$$

$$V_E = \underbrace{\left\{N\sum_i M_i(M_i - m_i)\frac{\sigma_i^2}{m_i}\right\}}_{\text{Within variance}} + \underbrace{\left\{\sum_i NY_i^2 - Y^2\right\}}_{\text{Between variance}},$$

where σ_i^2 is defined as for V_A.

Sampling system F:

$$\left[\frac{X_i + X_j}{(N-1)X},\ \left(\frac{Y_i' + Y_j'}{X_i + X_j}\right)X\right]$$

$$V_F = \underbrace{\left\{\sum_{i<j}\sum \frac{(Z_i + Z_j)X}{(N-1)(X_i + X_j)}\right\}}_{\text{Within variance}}$$

$$+ \underbrace{\left\{\sum_{i<j}\sum \frac{1}{(N-1)}\,\frac{(Y_i + Y_j)^2}{(X_i + X_j)}X - Y^2\right\}}_{\text{Between variance}},$$

where $Z_i = M_i(M_i - m_i)\sigma_i^2/m_i$ and σ_i^2 is defined as for V_A.

Sampling system G:

$$\left[\frac{F_{0i} + F_{0j}}{(N-1)F_0}, \left(\frac{Y_i' + Y_j'}{F_{0i} + F_{0j}}\right) F_0\right]$$

$$V_G = \underbrace{\left\{\sum_{i<j}\sum \frac{(Z_i + Z_j)F_0}{(N-1)(F_{0i} + F_{0j})}\right\}}_{\text{Within variance}}$$

$$+ \underbrace{\left\{\sum_{i<j}\sum \frac{1}{(N-1)} \frac{(Y_i + Y_j)^2}{(F_{0i} + F_{0j})} F_0 - Y^2\right\}}_{\text{Between variance}},$$

where Z_i, Z_j are defined as for V_F.

Sampling system H:

$$\left[\frac{2X_iX_j}{X^2}, \left(\frac{Y_i'}{X_i} + \frac{Y_j'}{X_j}\right) \frac{X}{2}\right]$$

$$V_H = \underbrace{\left\{\frac{1}{2}\left[X \sum_i \frac{M_i(M_i - m_i)}{X_i} \frac{\sigma_i^2}{m_i}\right]\right\}}_{\text{Within variance}}$$

$$+ \underbrace{\left\{\frac{1}{2}\left[X \sum_i \frac{Y_i^2}{X_i} - Y^2\right]\right\}}_{\text{Between variance}},$$

where σ_i^2 is defined as in V_A.

Sampling system K:

$$\left[\frac{2F_{0i}\cdot F_{0j}}{F_0^2}, \left(\frac{Y_i'}{F_{0i}} + \frac{Y_j'}{F_{0j}}\right) \frac{F_0}{2}\right]$$

$$V_K = \underbrace{\left\{\frac{1}{2}\left[F_0 \sum_i \frac{M_i(M_i - m_i)}{F_{0i}} \frac{\sigma_i^2}{m_i}\right]\right\}}_{\text{Within variance}}$$

$$+ \underbrace{\left\{\frac{1}{2}\left[F_0 \sum_i \frac{Y_i^2}{F_{0i}} - Y^2\right]\right\}}_{\text{Between variance}},$$

where σ_i^2 is defined as in V_A.

COMBINING INDEPENDENT TESTS OF SIGNIFICANCE*

ALLAN BIRNBAUM
Columbia University

It is shown that no single method of combining independent tests of significance is optimal in general, and hence that the kinds of tests to be combined should be considered in selecting a method of combination. A number of proposed methods of combination are applied to a particular common testing problem. It is shown that for such problems Fisher's method and a method proposed by Tippett have an optimal property.

1. THE PROBLEM AND SOME PROPOSED SOLUTIONS

THE problem of combining independent tests of significance has been discussed and illustrated by a number of writers, including Fisher [2], Karl Pearson (cf. [4]), Wallis [7], and E. S. Pearson [4], to which the reader is referred for general discussions to supplement the present brief section. The formal statistical problem may be stated as follows: A hypothesis H_0 is to be tested. An observed value t_1 of a statistic has been obtained; the best test of H_0 based on this statistic would indicate rejection at the u_1 significance level. That is, u_1 is the "probability level" corresponding to the observed value t_1; for example, if large values of the statistic are critical for H_0, then u_1 is the probability that a value as large as or larger than that observed will occur under H_0. Similarly, independent values of statistics, $t_2, \cdots, t_k$, have been obtained, and in the respective best tests of H_0 based on these statistics the corresponding "probability levels" are $u_2, \cdots, u_k$. The essential requirement of independence of the t_i's will be satisfied if each t_i is based on a separate and independent set of data; if each t_i is based on the same set of observations, the t_i's must be known to be statistically independent functions of the observations.

The problem of "combining these independent tests of significance" then is the problem of giving a test of H_0 on the basis of a set of observed values (probability levels) $u_1, u_2, \ldots, u_k$. The test is not to utilize the observed values $t_1, t_2, \ldots, t_k$; in general, it is assumed that

(a) either these values or else the forms of the distributions of $t_1, t_2, \ldots, t_k$, are unknown to the statistician confronted with the present problem; or

* Work sponsored by the Office of Naval Research.
The writer is grateful to Professor Henry Scheffé for helpful comments on the first draft of this paper. Responsibility for any remaining deficiencies is the writer's.

(b) this information is available but the distributions are such that there is no known or reasonably convenient method available for constructing a single appropriate test of H_0 based on $(t_1, t_2, \ldots, t_k)$.

Procedures for combining independent significance tests may be of practical use even in some situations in which the statistician has complete freedom to determine the design of a complex experiment. Suppose, for example, that a scientific hypothesis asserts that a change in the value of an independent experimental variable will alter the distribution of one or more of k observable variables $t_1, t_2, \ldots, t_k$. For example, an hypothesis to be tested may assert that administration to subjects of a certain drug will have at least one of the three effects:

(a) an increase in the mean of a certain measurable physiological quantity,

(b) an increase in the variance (within a subject) of a second measurable physiological quantity, and

(c) a decrease in the probability of a subject's correctly making a certain sensory discrimination.

Suppose that optimal tests for each of these effects separately could be based respectively on statistics t_1, t_2, and t_3. In such situations construction of a single optimal test for the presence of one or more of the effects may be difficult or impossible. However, combining statistically independent tests based on t_1, t_2, and t_3, a single test at a desired significance level can be given. With appropriate design, this test will also meet given requirements of power to detect one or more of the three effects. It is shown in Section 3 that for some problems such a test will even have certain efficiency properties.

To avoid technical complications not of direct interest here, let us assume that the t_i's have continuous distributions (densities). (See [7] for a discussion of the important discrete case.) Since u_i is the probability when H_0 is true of observing a value of our ith statistic at least as large as t_i, we may write

$$u_i = u_i(t_i) = \int_{t_i}^{\infty} p_i(t_i)dt_i \tag{1}$$

where $p_i(t_i)$ is the probability density function of t_i under H_0. Then the probability that u_i lies in any interval, say $u' \leqq u_i \leqq u''$, equals $u'' - u'$, or in other words u_i has a uniform distribution on the unit interval under H_0 with density

$$f(u_i) = \begin{cases} 1, & 0 \leqq u_i \leqq 1, \\ 0, & u_i < 0, u_i > 1, \end{cases} \tag{2}$$

for each i, and the u_i's are mutually independent.

Each method of combining tests is a rule prescribing that H_0 should be rejected whenever the set of values $(u_i, \cdots, u_k)$ falls in a certain critical region. Intuitively speaking, small values of the u_i's are indicative of rejection; to discuss satisfactorily the problem of constructing a critical region of values of $(u_i, \cdots, u_k)$, we must consider the possible distributions of the u_i's when H_0 is false. We shall assume here that whenever a u_i has a non-uniform distribution, it is distributed on the unit interval according to some (unknown) density function $g_i(u_i)$ which is non-increasing.[1]

Depending on the nature of the experimental situations in which the u_i's are obtained, the appropriate alternative hypothesis would be either:

H_A: All of the u_i's have the same (unknown) non-uniform, non-increasing density $g(u)$.

or:

H_B: One or more of the u_i's have (unknown) non-uniform densities $g_i(u_i)$.

Under H_A, the t_i's are statistics of the same kind obtained from k replications of an experiment, in which the underlying conditions are assumed to remain constant with H_0 false. Under H_B, the t_i's may be statistics of different kinds (for example, a normal mean and a normal variance), and the conditions under which the t_i's are obtained need not be the same; it is assumed only that H_0 is false in the case of at least one of the t_i's. H_A is seen to be a special case of H_B. Probably in the majority of applications, H_B is the appropriate alternative hypothesis.[2]

---

[1] This assumption is not a strong one for our purposes: Suppose large values of the statistic t are critical for testing H_0 against H_1, and the probability densities of t under H_0 and H_1, are $p(t)$ and $p'(t)$, respectively. Then the definition of the statistic u is

$$u = u(t) = \int_t^\infty p(t)dt, \tag{3}$$

so $du/dt = -p(t)$. If the probability density of t is $p'(t)$ then that of u is

$$g(u) = p'(t)/\left|du/dt\right| = p'(t)/p(t). \tag{4}$$

Hence $g(u)$ will be a non-increasing function of u if and only if $p'(t)/p(t)$ is a non-decreasing function of t. The latter condition is satisfied for most distributions commonly encountered in applied statistics, including those of normal, binomial, and Poisson means, normal variances, and all other distributions of the Koopman form described in Section 3 below.

[2] In some papers cited above, the distinction between the alternatives H_A and H_B seems not to have been made sufficiently clear. Problems corresponding to H_A are considered by Wallis on p. 238 of [7] and by Pearson on p. 142 of [4]. Problems corresponding to H_B are considered by Wallis on pp. 245–56 of [7] and by Pearson on p. 138 of [4].

Some of the methods which have been proposed for combining independent tests of significance (i.e., for constructing critical regions of values of $(u_1, u_2, \cdots, u_k)$) are the following:

(1) Fisher's [2] method: reject H_0 if and only if $u_1 \cdot u_2 \cdot \cdots \cdot u_k \leqq c$, where c is a predetermined constant corresponding to the desired significance level. Wallis, on pp. 231–34 of [7], discusses in detail Fisher's method of appropriately determining c. It turns out that $-2 \log u_1 \cdot u_2 \cdots u_k$ is distributed as chi-square with $2k$ degrees of freedom when H_0 is true. If d is such that

$$\text{Prob}\left\{\chi_{2k}^2 \geqq d\right\} = \alpha \tag{5}$$

where $1-\alpha$ is the desired significance level, then setting $-2 \log c = d$, we obtain $c = e^{-d/2}$.

(2) Karl Pearson's method: reject H_0 if and only if $(1-u_1)(1-u_2) \cdots (1-u_k) \geqq c$, where c is a predetermined constant corresponding to the desired significance level. In applications, c can be computed by a direct adaptation of the method used to calculate the c used in Fisher's method.

(3) Wilkinson's [8] methods: reject H_0 if and only if $u_i \leqq c$ for r or more of the u_i's, where r is a predetermined integer, $1 \leqq r \leqq k$, and c is a predetermined constant corresponding to the desired significance level. The k possible choices of r give k different procedures which we shall refer to as case 1 ($r=1$), case 2 ($r=2$), etc. For example, if $k=2$ and a test at the .95 significance level is desired, the case 1 procedure is: reject H_0 if either u_1 or u_2 or both equal or exceed $c=(.95)^{1/2}=.974$; the case 2 procedure is: reject H_0 if both u_1 and u_2 equal or exceed $c=1-(.05)^{1/2}=.776$. Case 1 was proposed earlier by Tippett [5].

In the following sections, certain bases for selecting methods of combination for particular problems will be developed.

2. A GENERAL CONDITION FOR ADMISSIBILITY OF METHODS OF COMBINATION

The following condition is readily seen to be satisfied by each of the proposed methods described above:

Condition 1: If H_0 is rejected for any given set of u_i's, then it will also be rejected for all sets of u_i^*'s such that $u_i^* \leqq u_i$ for each i.

Any method of combination which failed to satisfy this condition would seem unreasonable. In fact, it is not difficult to prove that the best test of H_0 against any particular alternative H_B of the kind described above satisfies Condition 1. (A proof is given in the Appendix.)

Since Condition 1 is satisfied by so many possible methods of combination, the question arises whether any further reasonable condition can be imposed to narrow still further the class of methods from which we must choose. The answer is no: So long as we consider the problem in the present generality, and not with reference to a particular kind of testing problem, there are no restrictions on the possible forms of the density functions $g_i(u_i)$ except that they be non-increasing. And it can be shown (see the Appendix) that for *each* method of combination satisfying Condition 1, we can find *some* alternative H_B represented by non-increasing functions $g_1(u_1), \cdots, g_k(u_k)$ against which that method of combination gives a best test of H_0.

These considerations prove that to find useful bases for choosing methods of combination, we must consider further the particular kinds of tests to be combined in any given problem. In the following sections, it is shown that certain methods are optimal for certain important categories of testing problems.

3. DISTRIBUTIONS OF THE KOOPMAN FORM

Nearly all of the density functions and discrete probability distribution functions encountered frequently in applied statistics can be written in the so-called Koopman form, which is

$$f(x, \theta) = c(\theta)a(\theta)^{t(x)}b(x) \tag{6}$$

where θ is a parameter of the distribution and x is an observed value, and a, b, c, and t denote arbitrary functions. Examples are

1. The binomial:

$$f(x, \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = (1-\theta)^n\left(\frac{\theta}{1-\theta}\right)^x\binom{n}{x}. \tag{7}$$

2. The normal, with known (say unit) variance and mean θ:

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2} = \frac{1}{\sqrt{2\pi}}e^{-\theta^2/2}e^{\theta x}e^{-x^2/2}. \tag{8}$$

Other examples are the Poisson and exponential distributions and the normal distribution with known mean and unknown variance.

Consider a problem of combining k independent significance tests, each of which is a test on a distribution of the Koopman form (all k distributions need not be of the same sort however; for example, one might be on a normal mean and another on a binomial mean, as in the illustration in Section 2 above). A method of combining these tests

will be equivalent to a test of a hypothesis specifying the values of k parameters,

$$H_0: \theta_1 = \theta_1^0, \cdots, \theta_k = \theta_k^0,$$

on the basis of the observed values of the statistics $t_1, \cdots, t_k$. For such problems, a minimal criterion for the reasonableness of a test is known. "Reasonableness" is used here in the sense of admissibility of a test, which may be defined as follows: A test is admissible if there is no other test with the same significance level which, without ever being less sensitive to possible alternative hypotheses, is more sensitive to at least one alternative. In other words, an admissible test is one which cannot be strictly (that is, uniformly) improved upon. A necessary condition for admissibility of a test of H_0 in our problem is that the acceptance region of the test (that is, the values of $(t_1, \cdots, t_k)$ for which the test accepts H_0) be convex. (A region is convex if the line segment connecting each pair of points in the region lies entirely in the region.) This is shown in [1].

We may illustrate both this condition for admissibility and its application to methods of combination by considering the problem of combining two tests on means of normal distributions with known (say unit) variances. (The performance of Fisher's method when applied to such a problem has been considered by Wallis (pp. 237–39 of [7]) and by Pearson (p. 142 of [4]). Let $\bar{x}_1$ denote the mean of a sample of n_1 observations obtained in an experiment in which the underlying population mean had the unknown value μ_1; let $\bar{x}_2$ be the mean of a sample of n_2 observations in a similar experiment in which the unknown population mean was μ_2. In this case any method of combining tests of the two hypotheses $\mu_1=0$ and $\mu_2=0$ is equivalent to a test of H_0: $\mu_1=\mu_2=0$; then H_A would specify $\mu_1=\mu_2\neq 0$; and H_B would specify that either μ_1 or μ_2 or both are not zero. Let $t_1=\sqrt{n_1}\bar{x}_1$ and $t_2=\sqrt{n_2}\bar{x}_2$. Then any method of combining the tests on μ_1 and μ_2 can be represented as a test of H_0 by its critical region in the (t_1, t_2) plane. Each of the methods of combination described above has been applied to the present problem, and the critical region corresponding to each method is illustrated in the figures below. The significance level $\alpha=0.5$ was used throughout. The tests on μ_1 and μ_2 to be combined were taken first to be against two-sided alternatives (Figures 1–4) and then against one-sided alternatives (Figures 6–9). In each case the critical region was obtained by first determining the values of u_1 and u_2 for which the method of combination considered would reject H_0 at the .05 significance level, and then plotting the corresponding values of t_1 and t_2 by

use of the equations relating the t_i's to the u_i's. These equations are, for the two-sided tests,

$$u_i = \frac{2}{\sqrt{2\pi}} \int_{|t_i|}^{\infty} e^{-v^2/2} dv, \qquad \text{for } i = 1, 2, \tag{9}$$

and for the one-sided tests,

$$u_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_i} e^{-v^2/2} dv, \qquad \text{for } i = 1, 2. \tag{10}$$

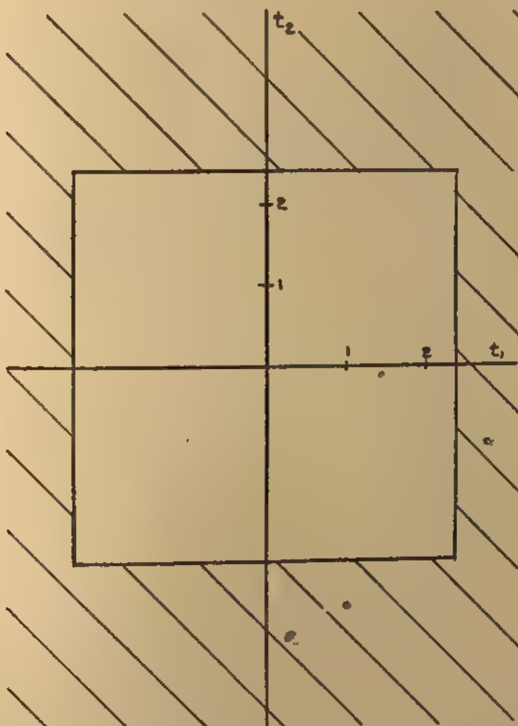


FIG. 1. Wilkinson's Method, Case 1.

We can now apply the condition for admissibility of a test described above: The acceptance regions obtained by Wilkinson's method, case 2, and by Pearson's method are not convex. Hence they represent tests of H_0, and corresponding methods of combination for the present problem, which can be strictly improved upon by other tests and corresponding methods of combination. Present knowledge does not provide

methods of finding tests which actually do strictly improve upon a given inadmissible test in problems like the present one. However, it seems advisable in selecting tests to restrict consideration to the class of admissible tests, and to select from this class a test which seems to have relatively good sensitivity (power) against the range of alternatives of interest.

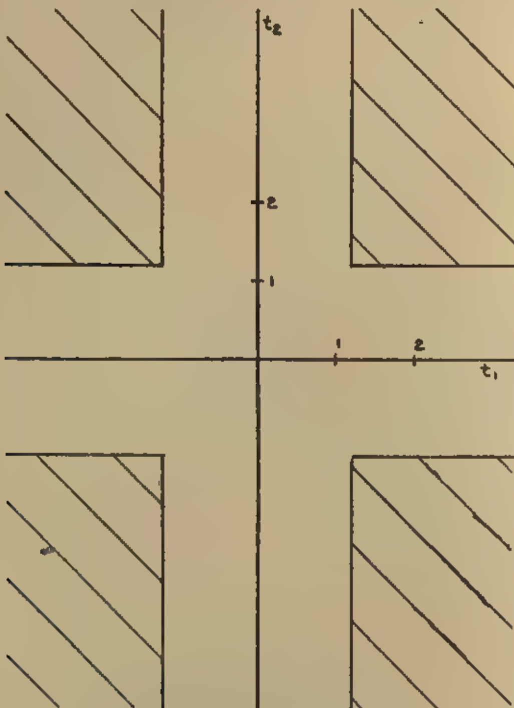


FIG. 2. Wilkinson's Method, Case 2.

It is shown in [1] that, for a category of problems including the present one, convexity of the acceptance region is a sufficient as well as necessary condition for admissibility.

The remaining two methods of combination, Wilkinson's case 1 and Fisher's, correspond to admissible tests. Inspection of Figures 1 and 3 suggests that each is fairly sensitive to departures from H_0 in all directions; that Fisher's method comes close to that test of H_0 (represented in Figure 5) which, if $n_1=n_2$ and if the seriousness of a departure from H_0 is measured by $\mu_1{}^2+\mu_2{}^2$, is the best test at the .05 level (as Wallis noted in [5]); and finally that Wilkinson's method, case 1,

gives a relative concentration of sensitivity to alternatives in which the departure from H_0 occurs in just one of the parameters. Similar observations can be made in Figures 6 and 8. Hence, it seems warranted to make a choice between the two methods remaining under consideration on the basis of a subjective appraisal of the context in which a problem like the present one actually occurs; probably in most cases Fisher's method would be preferred.

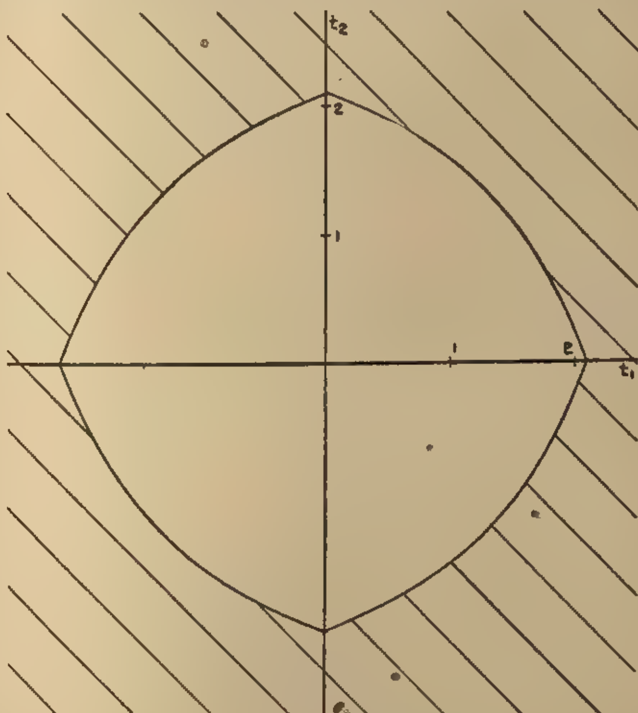


FIG. 3. Fisher's Method.

Having considered in detail a problem involving a particular distribution of the Koopman form, we proceed now to show that similar considerations apply to the whole class of such distributions. It can be verified easily that if Wilkinson's methods are used to combine tests on any such distributions, the result corresponds to a test whose acceptance region has a rectangular boundary like those in Figures 1, 2, 6, and 7, and is convex only in case 1. Hence, only case 1 of Wilkinson's method corresponds to an admissible test for certain of the Koopman-

form distributions being considered. The remaining cases of the method correspond to inadmissible tests for all Koopman-form distributions.

With little more difficulty it can be verified that Pearson's method does not give a test of H_0 with convex acceptance region for any Koopman-form distributions (consider the three points in the (t_1, t_2) plane corresponding to $(u_1, u_2)=(1-c, 0)$, to $(u_1, u_2)=(0, 1-c)$, and to

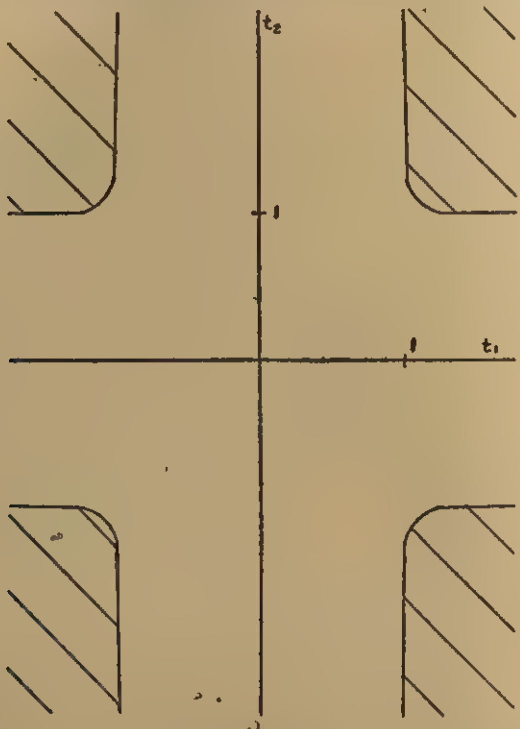


FIG. 4. Pearson's Method.

$(u_1, u_2)=(1-\sqrt{c}, 1-\sqrt{c})$, for the case $(k=2)$. Thus, Pearson's method also may be removed from consideration as inadmissible for Koopman-form distributions. Fisher's method does seem to give tests of H_0 with convex acceptance regions for Koopman-form distributions; consideration of the points in the (t_1, t_2) plane corresponding to $(u_1, u_2)=(1, c)$, to $(u_1, u_2)=(c, 1)$, and to $(u_1, u_2)=(\sqrt{c}, \sqrt{c})$ suggests this, and for particular distributions it may be possible to verify it fully without too much difficulty. For example, for

$$f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, \qquad x > 0, \tag{11}$$

to combine two one-sided tests on θ based on one observation each, we have

$$u_i = e^{-x_i/\theta_0}, \qquad \text{for } i = 1, 2, \tag{12}$$

and the critical region $u_1 u_2 \leqq c$ corresponds to a test with the convex acceptance region $x_1 + x_2 \leqq -\theta_0 \log c$.

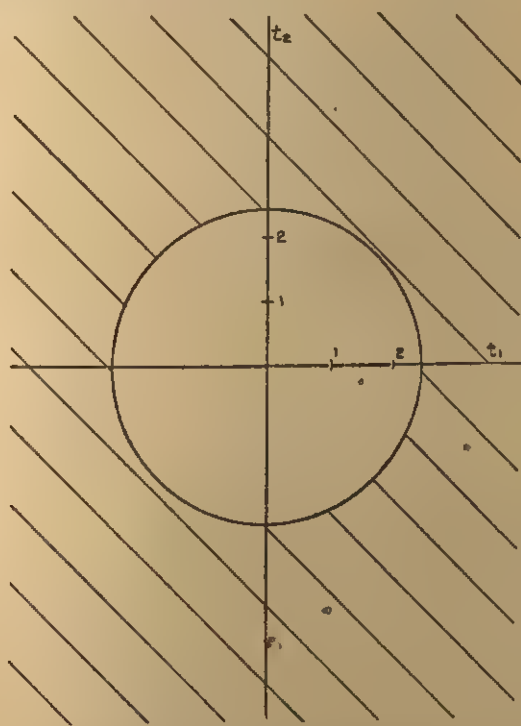


FIG. 5. Best Symmetric Test Against H_B.

4. CONCLUSIONS

While there is no single method of combining tests which is best for all problems, it appears that to combine independent tests on Koopman-form distributions (these include most distributions commonly occurring in applied statistics) one should choose between Fisher's

method and Wilkinson's method, case 1 Fisher's method appears to have somewhat more uniform sensitivity to the alternatives of interest in most problems. For any particular distributions, investigations may be made paralleling those above to obtain a still more conclusive basis for choice of a method of combination.

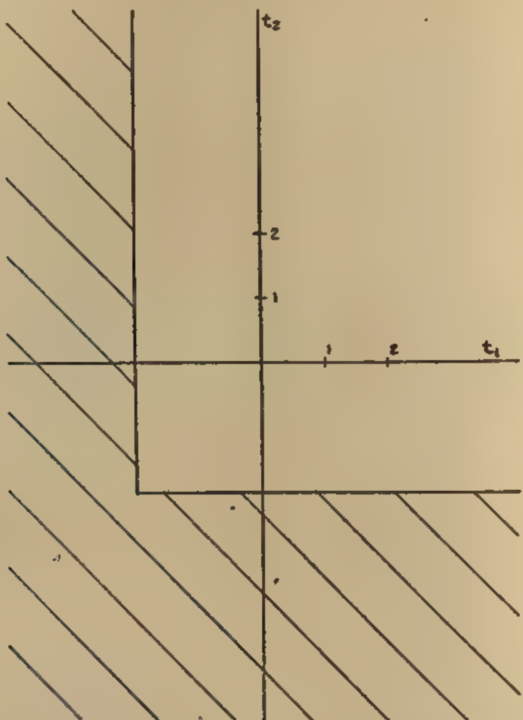


FIG. 6. Wilkinson's Method, Case 1. One-sided Alternatives.

APPENDIX

To prove that, as stated in Section 2 above, every test of H_0 which is best against some particular alternative specifying non-increasing densities, satisfies Condition 1, we use the well-known fact, proved for example in [3], that any best critical region consists of points satisfying

$$\lambda = \frac{g_1(u_1) \cdots g_k(u_k)}{f_1(u_1) \cdots f_k(u_k)} \geqq c, \qquad c \text{ some constant.} \tag{13}$$

Now $f_i(u_i)=1$ for $0\leqq u_i\leqq 1$, $i=1, \cdots, k$. Hence, $\lambda=g_1(u_2)\cdots g_k(u_k)$. As the $g_i(u_i)$'s are non-increasing, $g_1(u_1')\cdots g_k(u_k')\geqq g_1(u_1)\cdots g_k(u_k)\geqq c$ if $(u_1, \cdots u_k)$ is in the best critical region and if $u_i'\leqq u_i$ for $i=1, \cdots, k$. Thus Condition 1 is satisfied.

However, in general H_B (and even H_A) will include a whole set of possible forms of the $g_i(u_i)$'s, and it is not true in general that there will

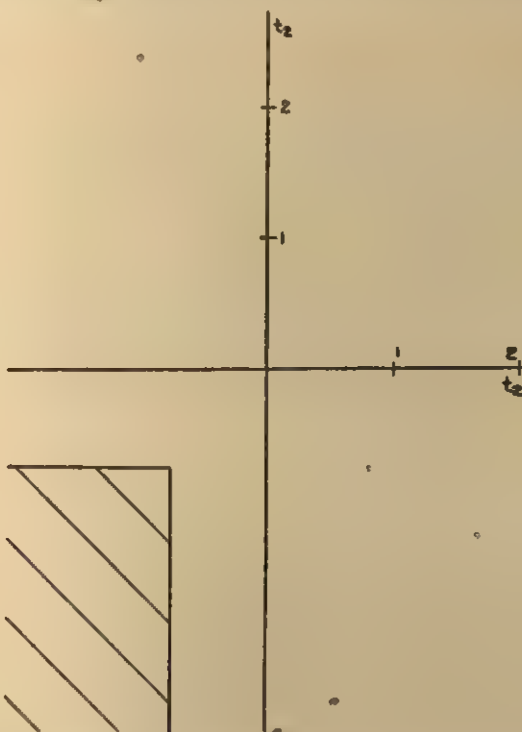


FIG. 7. Wilkinson's Method, Case 2. One-sided Alternatives.

exist a single test of H_0 which is uniformly best against all possibilities. This is illustrated most simply in the following case under H_B: If $g_1(u_1)$ is uniform and $g_2(u_2)$ is nonuniform, a best critical region consists of all (u_1, u_2) such that $u_2\leqq c$, c some constant; if $g_2(u_2)$ is uniform and $g_1(u_1)$ nonuniform, a best critical region consists of all (u_1, u_2) such that $u_1\leqq c'$, c' some constant; thus, there is not a single critical region which is best against each alternative. It can be verified directly that every best test of H_0 against a "Bayes mixture" of simple alternatives under

H_B also satisfies Condition 1. It follows, as shown by Wald in [6], that under general assumptions Condition 1 is a necessary condition for admissibility of a test of H_0 against a composite alternative H_B.

We shall show next that, as stated in Section 2 above, *each* method of combination meeting Condition 1 is *best* against *some* particular alternative hypothesis H_B. Taking $k=2$ for simplicity, any critical

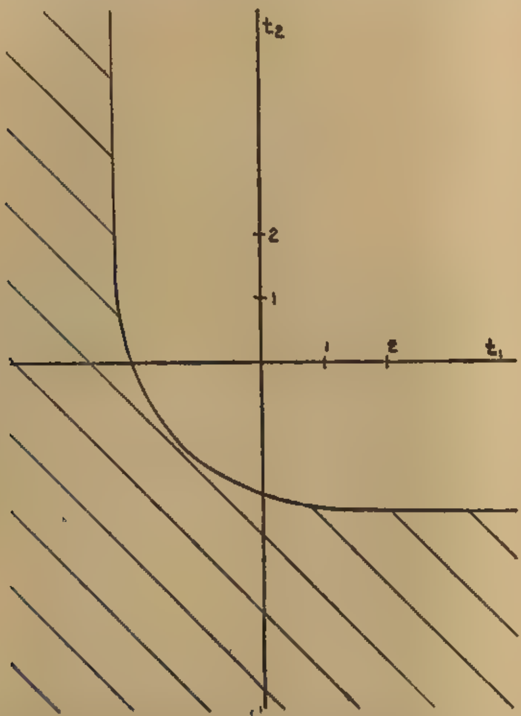


FIG. 8. Fisher's Method. One-sided Alternatives.

region w of values of (u_1, u_2), if it satisfies Condition 1, can be characterized by giving its boundary function $u_2(u_1)$, a non-increasing function such that w consists of all points (u_1, u_2) in the unit square $0 \leqq u_1 \leqq 1$, $0 \leqq u_2 \leqq 1$ for which $u_2 < u_2(u_1)$. Let $u_2(u_1)$ be any such boundary junction. Let $g_2(u_2) = \frac{2}{3}(2-u_2)$ for $0 \leqq u_2 \leqq 1$, and let $g_1(u_1) = \frac{3}{2}c(2-u_2(u_1))^{-1}$ for $0 \leqq u_1 \leqq 1$, where c is determined by the condition that $\int_0^1 g_1(u_1)du_1 = 1$. A best critical region for testing H_0 against the alternative $g_1(u_1)$, $g_2(u_2)$ is the set w' on which $g_1(u_1)g_2(u_2) > c$. But

$g_1(u_1)g_2(u_2) \equiv c(2-u_2)/(2-u_2(u_1)) > c$ if and only if $u_2 < u_2(u_1)$. Thus the arbitrarily given boundary function $u_2(u_1)$ characterizes a best critical region w'.

(Similar methods give analogous results for the problem of testing H_0 against H_A, with Condition 1 now strengthened by the requirement that the boundary function $u_2(u_1)$ be symmetric about the line $u_1 = u_2$.)

FIG. 9. Pearson's Method. One-sided Alternatives.

REFERENCES

[1] Birnbaum, Allan, "Characterizations of Complete Classes of Tests of Some Multiparametric Hypotheses, with Applications to Likelihood Ratio Tests," to be published in *Annals of Mathematical Statistics*.

[2] Fisher, R. A., *Statistical Methods for Research Workers*, Fourth and later Editions, Edinburgh and London, Oliver and Boyd, 1932 and later, Section 21.1.

[3] Neyman, J., *First Course in Probability and Statistics*, New York, Henry Holt and Company, 1950, Section 5.3.1, 304–8.

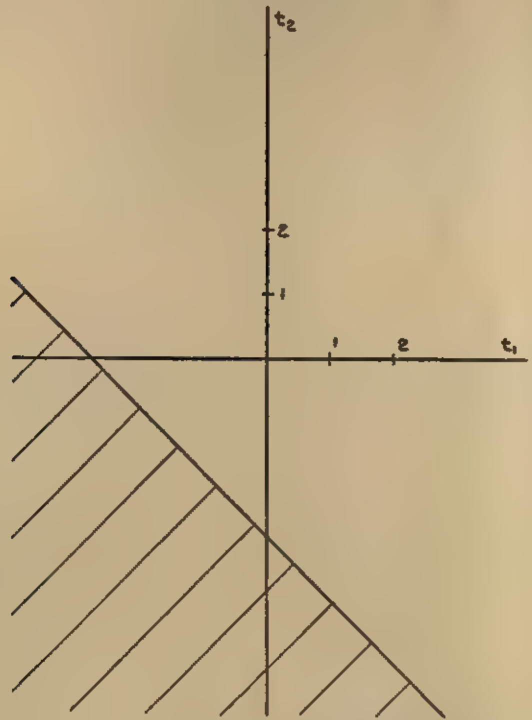


FIG. 10. Best Test of H_0 Against H_A. One-sided Alternatives.

[4] Pearson, E. S., "The Probability Integral Transformation for Testing Goodness of Fit and Combining Independent Tests of Significance," *Biometrika*, 30 (1938), 134–48.

[5] Tippett, L. H. C., *The Methods of Statistics*, First Edition, London, Williams and Norgate, Ltd., 1931, Section 3.5, 53–6.

[6] Wald, Abraham, *Statistical Decision Functions*, New York, J. Wiley and Sons, Inc., 1950, Theorem 3.20, 101.

[7] Wallis, W. Allen, "Compounding Probabilities from Independent Significance Tests," *Econometrica*, 10 (1942), 229–48.

[8] Wilkinson, B., "A Statistical Consideration in Psychological Research," *Psychological Bulletin*, 48 (1951) 156–7.

MINIMUM LIFE IN FATIGUE

A. M. Freudenthal and E. J. Gumbel
Columbia University

In conventional fatigue testing a specimen is repeatedly stressed in bending, torsion or tension-compression during imposed force-cycles of constant amplitude. The number N of cycles at which the specimen breaks, which is a function of the applied stress-amplitude S, is the observed variate [1]. The problem is to obtain, for specimens of specified material, shape and size, the probability of surviving up to N repetitions of a specified stress-cycle in a given testing procedure.

In a previous paper [2] fatigue was interpreted as an extremal phenomenon. A statistical scheme for the analysis of fatigue failures was developed and applied to the few observed series which satisfy the criteria for the applicability of any statistical procedure. The probability of surviving up to N cycles was obtained from the asymptotic theory of smallest values of a non-negative statistical variate. In this derivation it was assumed that a non-zero probability of failure existed even at the first cycle. This led to a good fit of the theory to tests made on copper and on aluminum at certain stress levels. However, it did not fit well enough the tests at other (lower) stress levels or of other metals. It appeared that such metals under certain stress levels would survive with certainty a substantial number of stress cycles, which thus represents a threshold value of the phenomenon.

Therefore a minimum number of cycles ("Minimum life") is now introduced into the survivorship function. Although this generalization does not change the fundamental nature of the probability function, a new estimate of the parameters becomes necessary. The extremal distribution function used in the following has already been introduced on a purely empirical basis by Weibull [6, 7] in his analysis of the distribution of the stress amplitude S for constant values of N.

1. THE LINEAR THEORY

In the previous theory [2] the probability $l(N)_S$ of surviving N cycles under the stress S was

$$(1.1) \qquad l(N)_S = \exp\left[-(N/V_S)^{\alpha_s}\right],$$

with the boundary conditions valid for all values of S

$$(1.1') \qquad l(0)_S = 1; \qquad l(\infty)_S = 0.$$

In equation (1.1) V_S is the characteristic number of cycles corresponding to the probability

$$l(V_S)_S = 1/e \tag{1.2}$$

and $1/\alpha_S$ is proportional to the standard deviation of the logarithms of the number of cycles. If we write

$$ln[-ln\, l(N)_S] = y; \qquad 1/\alpha_S' = 0.43429/\alpha_S \tag{1.3}$$

where *ln* stands for the natural logarithm, the relation (1.1) takes on the *linear form*

$$y = \alpha_S'(\log N - \log V_S); \qquad l(N)_S = \exp[-e^{+y}] \tag{1.4}$$

where y is a reduced variate without dimension and log stands for the common logarithm. The survivorship function (1.4) which is known by the actuaries as the Gompertz function has recently been tabulated by the National Bureau of Standards [3].

Instead of the theoretical limits $0 < N < \infty$ practical limits for the survivorship function are given by the interval

$$V_S e^{-12.25/\alpha_S} \leq N \leq V_S e^{2.50/\alpha_S} \tag{1.5}$$

for the number of cycles. In this approximation probabilities of the order 10^{-5} are neglected.

For homogeneous material and testing procedures, the probability of surviving a fixed number of cycles N decreases with increasing stress amplitude S. Since the relation (1.4) between log N and y is linear and α_S' is the slope of this line the α_S are constant and independent of S which means that the probabilities $l(N)_S$ traced on extremal probability paper against log N are parallel straight lines, which therefore cannot intersect. It follows that the estimates of α_S should be constant within errors of random sampling. However the acceptable domain of variation of the estimates cannot be established a priori since it depends upon the spacing of the stress levels, i.e., on the experimental design.

The first equation (1.4) gives a graphical criterion for the validity of the theory: The logarithms of the observed numbers of cycles N at fracture traced on extremal probability paper against the reduced variate y should be scattered about the straight line

$$\log N = \log V_S + y/\alpha_S'. \tag{1.6}$$

In this formula the y corresponding to the observed numbers $N_m (m = 1.2 \cdots n)$ are obtained from the plotting positions [2]

$$l(N_m)_S = 1 - m/(n + 1) \tag{1.7}$$

where m stands for the ranks of the observed numbers N_m arranged in increasing magnitudes and n is the total number of specimens tested.

It was shown [2] that the theory (1.1) fits torsion fatigue test results for copper and aluminum at relatively high stress levels. However, for low stress levels the test results approach a curve which, for small numbers of cycles, is definitely bending to the right (upwards). This means that fracture can occur only after a certain number of cycles. This number is henceforth called the *minimum life*. It is a physical constant for given material and testing procedure and its estimate is subject to statistical variations.

In addition to the graphical criterion, there is a numerical criterion: If the theory (1.1) holds, the arithmetic and geometric standard deviations $\sigma(N)$ and $\sigma(\log N)$ are related as shown in Table I of the previous paper [2]. This relation, traced in Figure 1 leads to the following

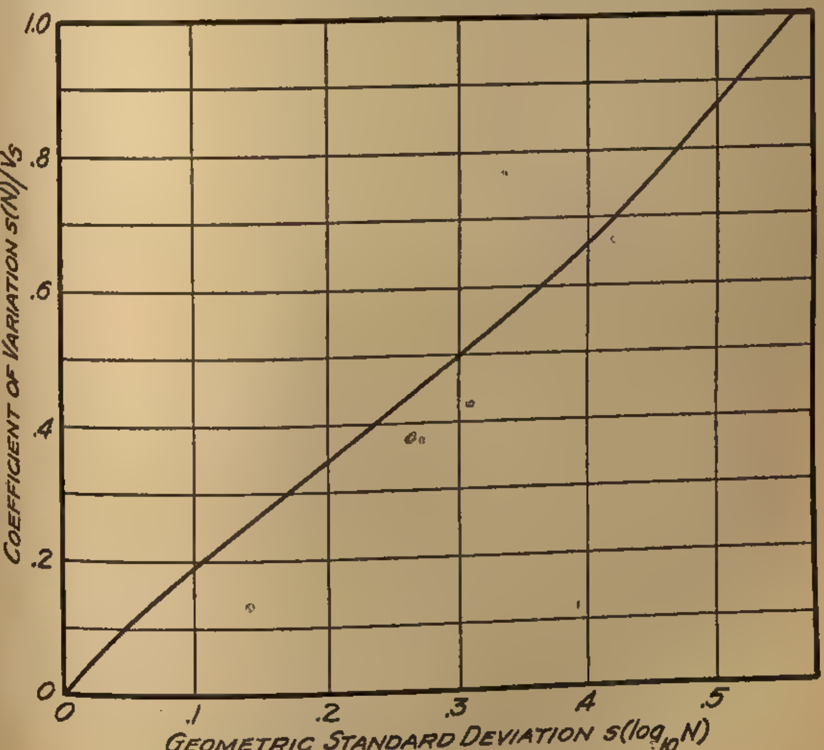


FIG. 1. Criterion for the validity of the linear theory.

procedure: We estimate the characteristic number V_S, calculate the standard deviation $s(N)$ and the geometric standard deviation $s(\log N)$, and check whether the quotient $s(N)/V_S$ corresponds to the value prescribed by the graph from the sample value $s(\log N)$. If this relation is not fulfilled, the assumption cannot be sustained.

2. THE GENERAL THEORY

If the assumption that the minimum life is zero is refuted by one of the above criteria, the asymptotic probability function (1.1) is used limiting the variate N by the condition $N \geqq N_{0,S}$, whence

$$l(N)_S = \exp\left[-\left(\frac{N - N_{0,S}}{V_S - N_{0,S}}\right)^{\alpha_S}\right] \tag{2.1}$$

with the properties

$$l(V_S)_S = 1/e \qquad l(N)_{0,S})_S = 1. \tag{2.2}$$

For a given stress amplitude S all specimens survive any number of cycles up to *the minimum life* $N_{0,S}$. For values approaching $N_{0,S}$ the survivorship function (2.1) is bent to the right and approaches a straight line perpendicular to the N axis as shown in Figures 7 and 9.

In (2.1) V_S and $N_{0,S}$ are parameters of location, have the dimension of N and the relation $V_S > N_{0,S} \geqq 0$ while $1/\alpha_S$ is a parameter of scale without dimension. In the case $N_{0,S}=0$ the formulae of the previous paper]2[are obtained. Within the framework of the extremal theory the generalization (2.1) of (1.1) is legitimate since a linear translation of an extreme is still an extreme.

The parameter V_S has the same meaning in the previous (linear) and the present (general) theory. Since it corresponds to a common fixed probability it decreases in both theories with increasing values of S. However, within the two theories different estimators have to be used. For all values of S where the minimum life $N_{0,S}$ differs from zero, it decreases with increasing S.

In the special case $\alpha_S=1$, $1/\alpha_S'=0.43429$ the probability of survival (2.1) degenerates into an exponential function. This probability traced on semi-logarithmic paper is a linear function of the numbers of cycles. The parameter V_S coincides with the mean $\bar{N}_S$. The standard deviation $\sigma(N)_S$ is

$$-\sigma(N)_S = V_S - N_{0,S}. \tag{2.3}$$

Therefore, the minimum life $N_{0,S}$ may be estimated in this case by the difference of the mean and the standard deviation of the number of

cycles. Another estimation based on the smallest number of cycles, which in this case is the most precise one, was given by J. Neyman and E. S. Pearson [4]. This model seems unrealistic for fatigue observation because the exponential function implies that the expectation of future life is independent of the preceding number of cycles, i.e., of the "history." This assumption is compatible with certain physical processes, such as radio-active decay, but not with fatigue. In fact all observations available on fatigue lead to estimations for α_S which exceed unity.

For any value of α_S the median life $\overline{N}_S$ obtained from (2.1) is

$$\overline{N}_S - N_{0,S} = (V_S - N_{0,S})(\ln 2)^{1/\alpha_S}. \tag{2.4}$$

To obtain the modal life $\tilde{N}_S$ consider the distribution $p(N)_S$ of the number of cycles at failure obtained from (2.1) as

$$p(N)_S = \frac{\alpha_S}{V_S - N_{0,S}} \left(\frac{N - N_{0,S}}{V_S - N_{0,S}} \right)^{\alpha_S - 1} \exp \left[- \left(\frac{N - N_{0,S}}{V_S - N_{0,S}} \right)^{\alpha_S} \right]. \tag{2.5}$$

Differentiation with regard to N leads to the mode

$$\tilde{N}_S - N_{0,S} = (V_S - N_{0,S})(1 - 1/\alpha_S)^{1/\alpha_S}. \tag{2.6}$$

A mode exists only for $1/\alpha_S < 1$; $1/\alpha_S' < 0.43429$ and

$$\text{the mode} \begin{Bmatrix} \text{precedes} \\ \text{equals} \\ \text{exceeds} \end{Bmatrix} \text{the median if} \tag{2.7}$$

$$1/\alpha_S \begin{bmatrix} > \\ = 0.30685 \\ < \end{bmatrix}; \quad 1/\alpha_S' \begin{bmatrix} > \\ = 0.13326 \\ < \end{bmatrix}.$$

For $\alpha_S = 3.25889$ the distribution (2.5) is nearly symmetrical. Three other pseudosymmetrical cases will show up later.

The density of probability at the mode increases with α_S. This is shown in Figure 2 where $(N - N_{0,S})/(V_S - N_{0,S})$ is used as abscissa. For constant values of $N_{0,S}$ the median $\overline{N}_S$ and the mode $\tilde{N}_S$ converge with increasing values of α_S towards the characteristic number V_S. It will be shown in paragraph 3 that the same holds for the mean $\overline{\overline{N}}_S$.

The probabilities of survival $1/2$ and $1/e$ at the median $\overline{N}_S$ and at V_S are fixed. The probability $l(N)_S$ at the mode depends upon α_S. The median and the mode depend upon the three parameters V_S, $N_{0,S}$, α_S and the same will be shown to hold for the mean, while the number of cycles V_S is itself one of the parameters of the distribtuion. Therefore,

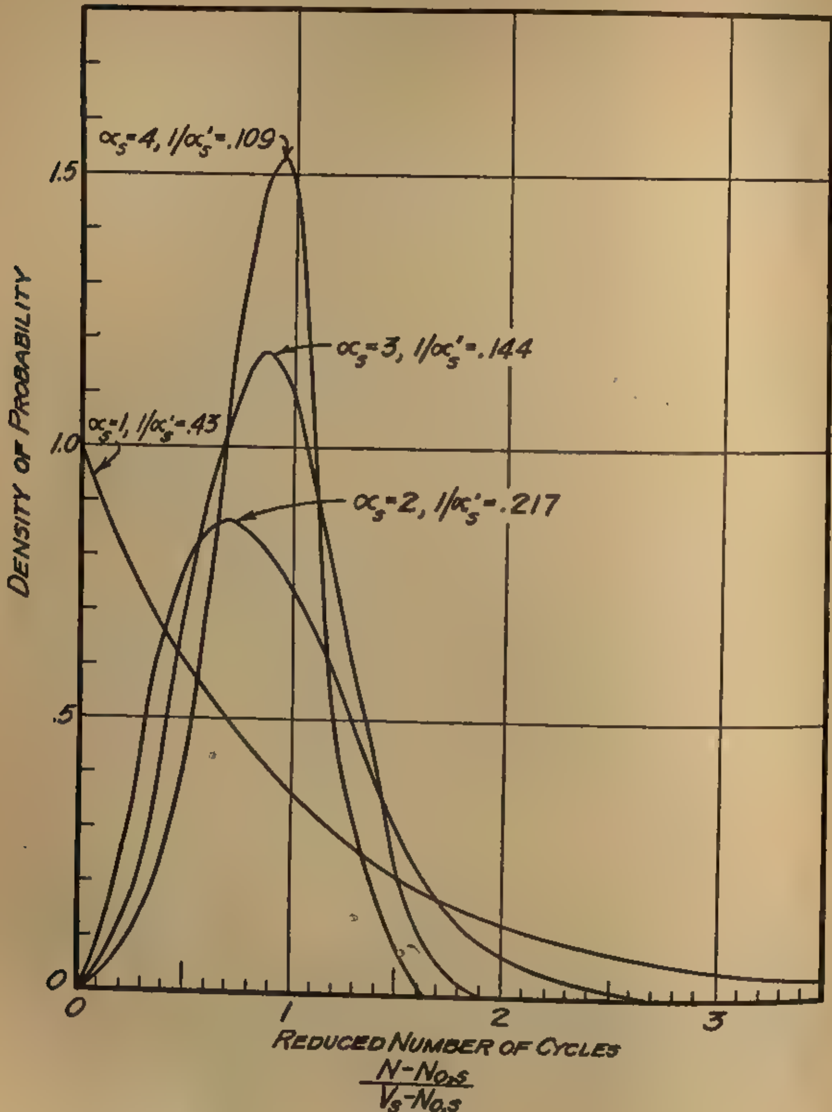


FIG. 2. Influence of α_S on the shape of the distributions of N_S.

an estimate of the location parameter V_S is more convenient to characterize the fatigue failure of a given material than the mean, the mode, or the median.

For the graphical representation of the numbers N at fracture the

reduced variate y defined in (1.3) is used. Equation (2.1) then becomes, in analogy to (1.6),

$$\log(N - N_{0,S}) = \log(V_S - N_{0,S}) + y/\alpha_S', \tag{2.8}$$

where α_S' and α_S are related by (1.3). In the previous (linear) theory the parameter $1/\alpha_S'$ is the slope of $\log N$ plotted against y. In the present theory it is the slope of $\log(N - N_{0,S})$ against y. Within the previous theory α_S is independent of S, within the present theory it may depend upon S. If V_S is large compared to $N_{0,S}$ the logarithms of the number of cycles at fracture traced on the extremal probability paper against y are practically linear as long as N is in the neighborhood of V_S. However, if $N_{0,S}$ is of the same order as V_S the curve is bent to the right (upwards) for $N < V_S$. For the same values of V_S and $N_{0,S}$, the asymptotic value $N_{0,S}$ is approached more quickly if $1/\alpha_S$ becomes larger. Two curves are parallel if they have the same value of $1/\alpha_S$ and $(V_S - N_{0,S})$ although the parameters V_S and $N_{0,S}$ may differ.

The limiting condition (1.5) leads from (2.8) to a number of cycles

$$N_{\omega,S} = N_{0,S} + (V_S - N_{0,S})e^{2.5/\alpha_S} \tag{2.9}$$

for which $l(N_\omega)_S$ is practically zero.

For a large number of cycles and small probabilities of survival a relatively small increase in the number of cycles considerably reduces the probability of survival. This holds for the linear and the general theory and corresponds to the popular statement of the straw that broke the camel's back. For high probabilities of survival a considerable decrease in the number of cycles is necessary in the linear theory in order to increase the probability of survival by a small amount, while in the general theory a small decrease in the number of cycles has a large influence on the probability of survival.

It has often been assumed that the logarithms of the number of cycles to fracture in fatigue are normally distributed. However, in this case, the probability of survival converges in the same way to zero and to unity. Therefore the use of this distribution for extrapolation does not seem to be legitimate.

3. ESTIMATE OF THE PARAMETERS

Since the relation (2.8) between $\log N$ and y is no longer linear as was the case for $N_{0,S} = 0$ the graphical estimate of the parameters given previously [2] is no longer feasible. Instead the classical method of moments advocated by Weibull [7] for the analysis of breaking strengths is used. The special case $\alpha_S = 1$ was settled in (2.3).

In the general case $\alpha_S \neq 1$ the reduced moments of order k obtained from the distribution (2.5) are the gamma functions

$$\left(\frac{N - N_{0,S}}{V_S - N_{0,S}}\right)^k = \Gamma(1 + k/\alpha_S). \tag{3.1}$$

For $k=1, 2, 3$ three equations are obtained which may be used for the estimation of the three parameters. The mean,

$$\overline{N}_S - N_{0,S} = (V_S - N_{0,S})\Gamma(1 + 1/\alpha_S), \tag{3.2}$$

depends upon the three parameters and the probability at the mean depends upon α_S. The relation between the mean and the characteristic number of cycles is

$$\overline{N}_S \gtrless V_S \text{ if } 1/\alpha_S \gtrless 1; \qquad 1/\alpha_S' \gtrless 0.43429.$$

For increasing values of α_S the mean converges to the characteristic number V_S. The variance $\sigma^2(N)_S$ obtained from (3.1) and 3.2),

$$\sigma^2(N)_S = (V_S - N_{0,S})^2[\Gamma(1 + 2/\alpha_S) - \Gamma^2(1 + 1/\alpha_S)], \tag{3.3}$$

also depends upon the 3 parameters. The corresponding sample variance $s^2(N)_S = s_S^2$ is obtained by the usual procedure as

$$s_S^2 = n(\overline{N_S^2} - \overline{N}_S^2)/(n - 1). \tag{3.3'}$$

The skewness $\sqrt{\beta_{1,S}}$ defined by

$$\sqrt{\beta_{1,S}} = \mu_{3,S}\sigma_S^{-3} \tag{3.4}$$

where $\mu_{3,S}$ is the third central moment is

$$\sqrt{\beta_{1,S}} = [\Gamma(1+3/\alpha_S) - 3\Gamma(1+2/\alpha_S)\Gamma(1+1/\alpha_S) + 2\Gamma^3(1+1/\alpha_S)] \\ [\Gamma(1+2/\alpha_S) - \Gamma^2(1+1/\alpha_S)]^{-3/2}. \tag{3.5}$$

and depends only upon the parameter $1/\alpha_S$. If the population value $\sqrt{\beta_{1,S}}$ is replaced by the sample value

$$\sqrt{b_{1,S}} = \frac{\sqrt{n(n-1)}}{n-2} \frac{\overline{N_S^3} - 3\overline{N_S^2}N_S + 2\overline{N}_S^3}{(\overline{N_S^2} - \overline{N}_S^2)^{3/2}}, \tag{3.5'}$$

an estimate of $1/\alpha_S$ is obtained. To facilitate this procedure, the right side of equation (3.5) as a function of $1/\alpha_S$ is given in Table 1, cols. 4 and 1.[1] The value of $1/\alpha_S'$ in equation (2.8) is obtained then from (1.3).

The two remaining parameters of location, the characteristic number

[1] This table was calculated by Gladys R. Garabedian of Stanford University. The authors take this occasion to thank her for this important contribution.

V_S and the minimum life $N_{0,S}$, are simple to estimate. Combination of (3.3) and (3.2) leads to

$$V_S = \overline{N}_S + \sigma_S A(\alpha_S), \tag{3.6}$$

where the standardized distance from the characteristic number to the mean

$$A(\alpha_S) = [1 - \Gamma(1 + 1/\alpha_S)][\Gamma(1 + 2/\alpha_S) - \Gamma^2(1 + 1/\alpha_S)]^{-1/2} \tag{3.6'}$$

is given in Table 1, col. 3.

Since $1/\alpha_S$ is estimated from (3.5), the parameter V_S may be estimated from (3.6) after replacing the population mean and standard deviation by the sample values. The result can be checked from the observations traced on the extremal probability paper with the help of the first equation (2.2).

To estimate the minimum life the value of $\overline{N}_S$ given in (3.6) is introduced in (3.2) whence

$$N_{0,S} = V_S - \sigma_S B(\alpha_S) \tag{3.7}$$

where the standardized distance from the characteristic number to the minimum life

$$B(\alpha_S) = [\Gamma(1 + 2/\alpha_S) - \Gamma^2(1 + 1/\alpha_S)]^{-1/2} \tag{3.7'}$$

is given in Table 1, col. 2. For the estimation of $N_{0,S}$ we use the previous estimates of $1/\alpha_S$ and of V_S and replace the population value σ_S by the sample value s_S.

Thus the estimate $\hat{\alpha}_S$ is obtained from $\sqrt{b_{1,S}}$, equation (3.5′), with the help of Table 1. The two other estimates are from (3.6) and (3.7),

$$\widehat{V}_S = N_S + s_S A(\widehat{\alpha}_S); \qquad N_{0,S}^{*} = V_S - s_S B(\widehat{\alpha}_S). \tag{3.8}$$

The minimum life is thus estimated directly, without using iterated procedures.

The result may be checked by another estimate based on the observed smallest number N_1 of cycles at fracture. Its plotting position $1-1/(n+1)$ for n observations obtained from (1.7) and equation (2.1) lead to the unbiased estimate

$$N'_{0,s} = \frac{N_1(n + 1)^{1/\widehat{\alpha}_s} - \widehat{V}_s}{(n + 1)^{1/\widehat{\alpha}_s} - 1}. \tag{3.8'}$$

This estimate is always smaller than the observed smallest number of cycles. Of course, this method also requires the previous estimate of the two other parameters $1/\alpha_S$ and V_S.

TABLE 1

ESTIMATION OF THE THREE PARAMETERS

1 Scale parameter $1/\alpha_S$	2 Multiple of standard deviation for minimum life $N_{0,S}$ $B(\alpha_S)$ equ. (3.7′)	3 Multiple of standard deviation for parameter V_S $A(\alpha_S)$ equ. (3.6′)	4 Reduced 3d moment $\sqrt{\beta_1(\alpha_S)}$ equ. (3.5)
.01	78.9817	.4481	−1.0813
.02	39.9890	.4461	−1.0249
.03	26.9862	.4439	− .9707
.04	20.4808	.4416	− .9185
.05	16.5744	.4392	− .8680
.06	13.9673	.4366	− .8191
.07	12.1029	.4339	− .7717
.08	10.7024	.4310	− .7258
.09	9.6114	.4281	− .6811
.10	8.7369	.4250	− .6376
.11	8.0199	.4219	− .5953
.12	7.4209	.4186	− .5540
.13	6.9128	.4152	− .5137
.14	6.4761	.4118	− .4743
.15	6.0965	.4082	− .4357
.16	5.7633	.4046	− .3980
.17	5.4682	.4008	− .3610
.18	5.2050	.3970	− .3247
.19	4.9686	.3931	− .2891
.20	4.7549	.3891	− .2541
.21	4.5608	.3850	− .2197
.22	4.3835	.3809	− .1858
.23	4.2209	.3767	− .1525
.24	4.0711	.3724	− .1196
.25	3.9326	.3681	− .0872
.26	3.8041	.3637	− .0553
.27	3.6844	.3592	− .0237
.28	3.5727	.3547	+ .0075
.29	3.4680	.3501	.0383
.30	3.3698	.3455	.0687
.31	3.2774	.3408	.0989
.32	3.1901	.3361	.1287
.33	3.1077	.3313	.1583
.34	3.0296	.3265	.1876
.35	2.9554	.3217	.2167

1	2	3	4
	Multiple of standard deviation		
Scale parameter $1/\alpha_S$	for minimum life $N_{0,S}$ $B(\alpha_S)$ equ. (3.7′)	for parameter V_S $A(\alpha_S)$ equ. (3.6′)	Reduced 3d moment $\sqrt{\beta_1(\alpha_S)}$ equ. (3.5)
.36	2.8849	.3168	.2455
.37	2.8178	.3119	.2741
.38	2.7537	.3069	.3024
.39	2.6925	.3019	.3306
.40	2.6339	.2969	.3586
.41	2.5778	.2919	.3865
.42	2.5239	.2868	.4141
.43	2.4721	.2817	.4417
.44	2.4224	.2766	.4691
.45	2.3755	.2715	.4963
.46	2.3282	.2663	.5235
.47	2.2836	.2612	.5505
.48	2.2406	.2560	.5775
.49	2.1989	.2508	.6043
.50	2.1587	.2456	.6311
.51	2.1196	.2404	.6578
.52	2.0818	.2352	.6845
.53	2.0451	.2299	.7110
.54	2.0095	.2247	.7376
.55	1.9749	.2195	.7640
.56	1.9412	.2142	.7905
.57	1.9085	.2090	.8169
.58	1.8767	.2038	.8433
.59	1.8456	.1985	.8697
.60	1.8154	.1933	.8960
.61	1.7859	.1881	.9224
.62	1.7571	.1829	.9488
.63	1.7290	.1777	.9751
.64	1.7016	.1725	1.0015
.65	1.6748	.1673	1.0279
.66	1.6486	.1621	1.0544
.67	1.6230	.1570	1.0808
.68	1.5870	.1518	1.1073
.69	1.5734	.1467	1.1338
.70	1.5494	.1416	1.1604
.71	1.5259	.1365	1.1870
.72	1.5029	.1314	1.2137

1	2	3	4
	Multiple of standard deviation		
Scale parameter $1/\alpha_S$	for minimum life $N_{0.S}$ $B(\alpha_S)$ equ. (3.7′)	for parameter V_S $A(\alpha_S)$ equ. (3.6′)	Reduced 3d moment $\sqrt{\beta_1(\alpha_S)}$ equ. (3.5)
.73	1.4803	.1263	1.2404
.74	1.4581	.1213	1.2672
.75	1.4364	.1163	1.2941
.76	1.4151	.1113	1.3210
.77	1.3942	.1063	1.3480
.78	1.3737	.1013	1.3751
.79	1.3535	.0964	1.4023
.80	1.3338	.0915	1.4295
.81	1.3443	.0866	1.4569
.82	1.2952	.0818	1.4844
.83	1.2765	.0770	1.5119
.84	1.2580	.0722	1.5396
.85	1.2399	.0674	1.5674
.86	1.2220	.0627	1.5953
.87	1.2045	.0580	1.6233
.88	1.1872	.0533	1.6514
.89	1.1703	.0487	1.6797
.90	1.1536	.0441	1.7080
.91	1.1371	.0395	1.7366
.92	1.1209	.0350	1.7652
.93	1.1050	.0305	1.7940
.94	1.0893	.0260	1.8230
.95	1.0738	.0216	1.8521
.96	1.0586	.0172	1.8814
.97	1.0436	.0129	1.9108
.98	1.0289	.0085	1.9403
.99	1.0143	.0042	1.9701
1.00	1.0	0.0	2.0
1.0	1.	0.	2.
1.2	.7522	− .0766	2.6400
1.4	.5633	− .1364	3.3820
1.6	.4184	− .1797	4.2621
1.8	.3076	− .2081	5.3235
2.	.2236	− .2236	6.6188
3.	.03824	− .1912	19.5849
4.	.00502	− .1154	60.0917
5.	.00053	− .0626	190.1132

Finally, the theoretical numbers N corresponding to given probabilities $l(N)_S$ are obtained from (2.8) as

$$N = \widehat{N}_{0,S} + (\widehat{V}_S - \widehat{N}_{0,S})e^{y/\widehat{\alpha}_s}. \tag{3.9}$$

4. INFLUENCE OF THE PARAMETERS

To show the influence of the parameters the values $N_0 \equiv N_{0,S} = 1{,}000$; $1/\alpha' \equiv 1/\alpha_S' = 0.1$ and different values of V_S are chosen. In this simpli-

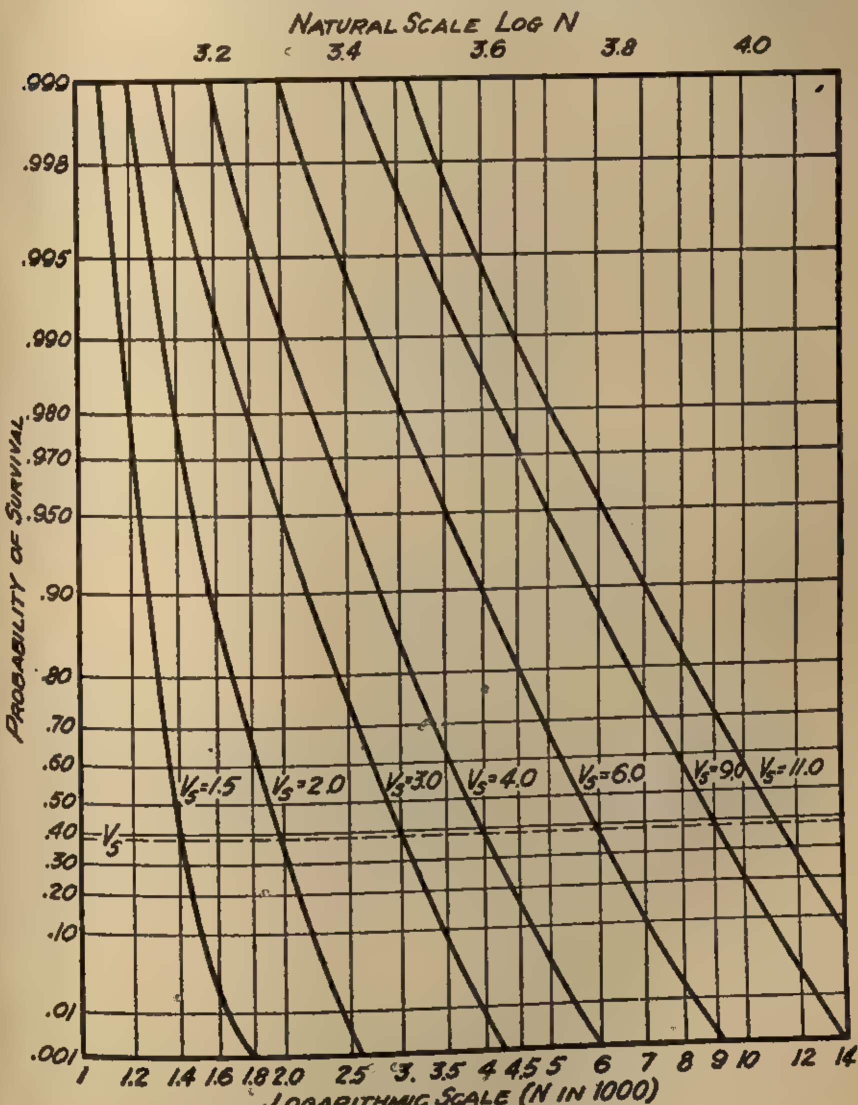


FIG. 3. Theoretical survivorship functions $l(N)_S$ for constant $N_{0,S} = 1.0$, $1/._S' = 0.1$ and various values of V_S.

fied scheme it is assumed that the parameter α and the minimum life N_0 are both independent of S, while α_S may and $N_{0.S}$ will depend upon S. The numbers of cycles at fracture N in 1,000 obtained from (2.8)

$$\log (N-1) = \log (V_S - 1) + 0.1\, y \tag{4.1}$$

are traced in Figure 3. The graph shows how the survivorship functions for different values of V_S converge to unity for a common value of N_0. Within the observable part $0.0476 \leqq l(N)_S \leqq 0.9524$ for $n=20$ specimens obtained from (1.7) the survivorship functions look fairly linear and the slopes seem to increase systematically with decreasing V_S, i.e., with increasing stress levels, while in reality it was assumed that the minimum life is invariant against changes in S and that the parameters $1/\alpha_S$ are constant. Thus the graph may serve as a warning against relying too much on the graphical representation for the usual small samples.

Since the normal distribution has sometimes been used in connection with fatigue observations, it is worthwhile to analyze under what conditions the distributions (2.5) *look* symmetrical. In addition to the pseudo-symmetrical case (2.7) where the median and the mode coincide three other pseudo-symmetrical cases exist. Comparison of equations (2.4), (2.6), and (3.2) shows that the mean is equal to the median

$$\overline{N}_S = \check{N}_S \text{ if } 1/\alpha_S = 0.29075, \tag{4.2}$$

and that the mean is equal to the mode

$$\overline{N}_S = \tilde{N}_S \text{ if } 1/\alpha_S = 0.30189. \tag{4.3}$$

Table 1 shows the existence of a fourth pseudo-symmetrical case. The third central moment is zero for $1/\alpha_S = 0.27760$. It follows from Table 1 that the distributions look symmetrical if the skewness is near to the interval

$$0 \leqq \sqrt{\beta_{1.S}} \leqq 0.1. \tag{4.4}$$

Then the scale parameter is near to the interval

$$0.27 < 1/\alpha_S < 0.31; \quad 3.2 < \alpha_S < 3.7; \quad 0.12 < 1/\alpha_S' < 0.13. \tag{4.5}$$

The four values of α_S are so close one to another that no practical distinction between these cases is possible. Of course none implies a normal distribution of the number of cycles at fracture.

The values of Table 1 are drawn in Figures 4, 5, and 6. Figure 4 shows the parameter $1/\alpha_S$ and $1/\alpha_S'$ and the standardized distances from the characteristic number to the mean and to the minimum life

(4.6) $A(\alpha_S) = (V_S - \overline{N}_S)/\sigma_S; \qquad B(\alpha_S) = (V_S - N_{0,S})/\sigma_S$

as functions of the skewness $\sqrt{\beta_{1,S}}$. The distances decrease with increasing skewness. Within the interesting domain of $1/\alpha_S$ the distance from the mean to V_S is smaller than the distance from $N_{0,S}$ to the mean. Figure 5 compares the standardized distance from V_S to $N_{0,S}$ to the standardized distance from $\overline{N}_S$ to $N_{0,S}$. Both are traced as func-

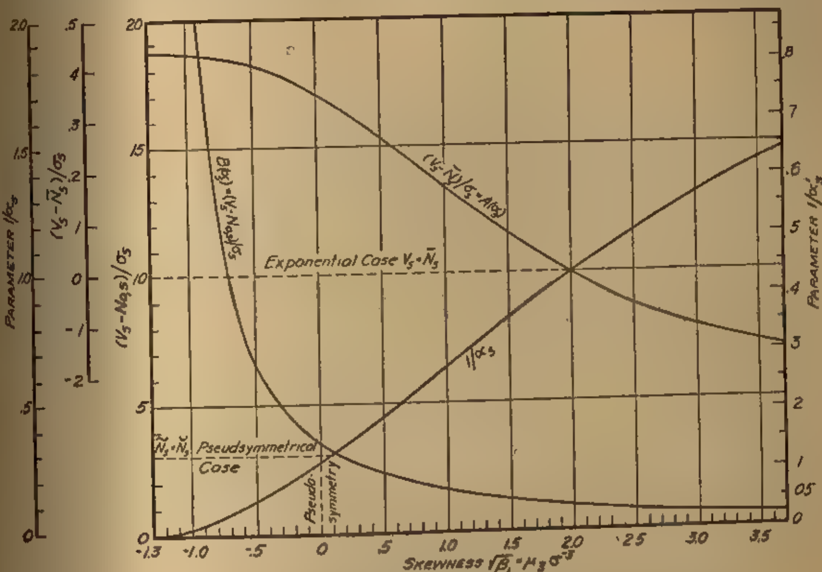


FIG. 4. Parameter $1/\alpha_S$ and standardized distances $A(\alpha_S)$ and $B(\alpha_S)$ as functions of skewness. (See Table 1.)

tions of the skewness and of the parameter $1/\alpha_S$ and $1/\alpha_S'$. Finally Figure 6 which shows the standardized distance from V_S to $N_{0,S}$ as function of $1/\alpha_S$ and $1/\alpha_S'$ facilitates the estimation of the minimum life.

The estimations $N_{0,S}$, V_S, and $1/\alpha_S$ and the three statistics N_S, s_S, $\sqrt{b_{1,S}}$ are related by the sample analogs of the three equations (3.5), (3.6), and (3.7) which will now be analyzed. The estimation of the parameter $1/\alpha_S$ depends only on the statistic $\sqrt{b_{1,S}}$. The population value $1/\alpha_S$ increases with the skewness as shown in Figure 4. In the previous (linear) theory this parameter was a function of the standard deviation of the logarithms of the number of cycles. In the present (general) theory it is a function of the skewness of the number of cycles.

The partial derivatives of V_S and $N_{0,S}$ with respect to each of the

three statistics $\overline{N}_S$, s_S, $\sqrt{b_{1,S}}$, obtained from (3.8) lead to the following relations: For constant values of the standard deviation and the skewness the estimates of the characteristic number of cycles $\hat{V}_S$ and of the

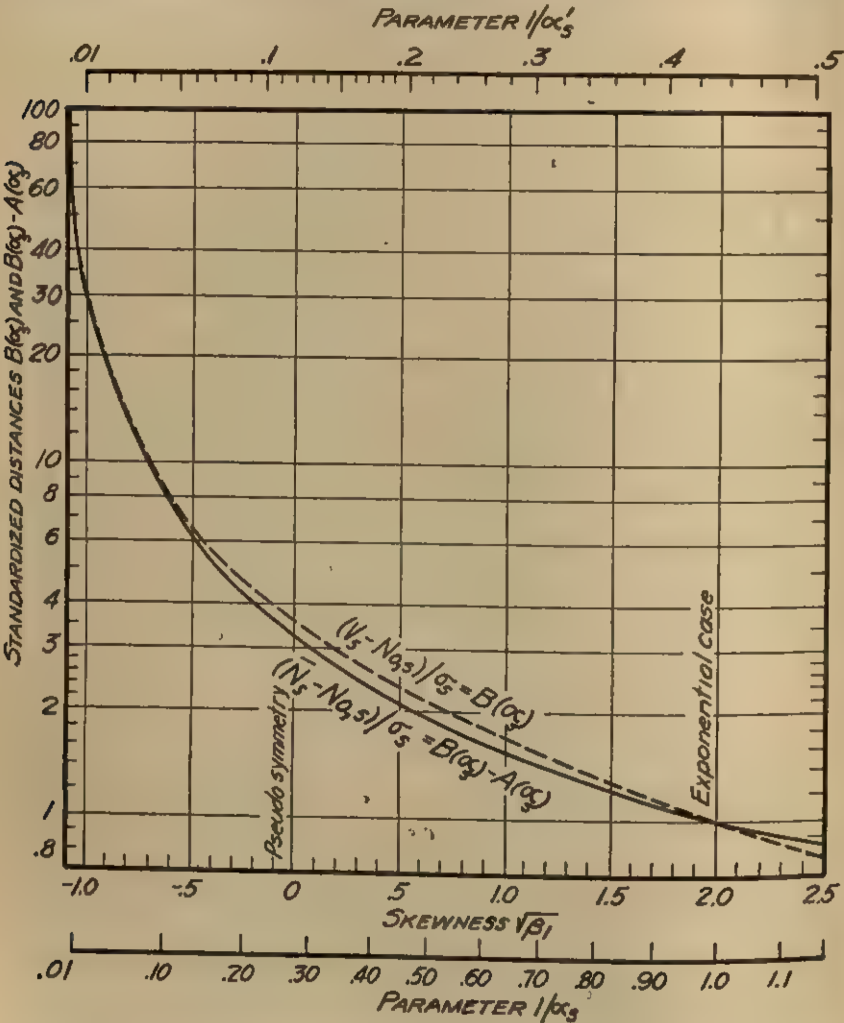


FIG. 5. Standardized distances as functions of parameter $1/\alpha_S$. (See Table 1.)

minimum life $\hat{N}_{0,S}$ increase proportionally to the mean $\overline{N}_S$. For constant mean and skewness and increasing standard deviations the estimates of the characteristic number of cycles increase for $1/\alpha_S<1$ and decrease for $1/\alpha_S>1$ and the minimum life decreases. For constant

mean and standard deviation the estimates of the characteristic number decrease and the minimum life increases with the skewness. For constant mean and skewness and increasing standard deviation the

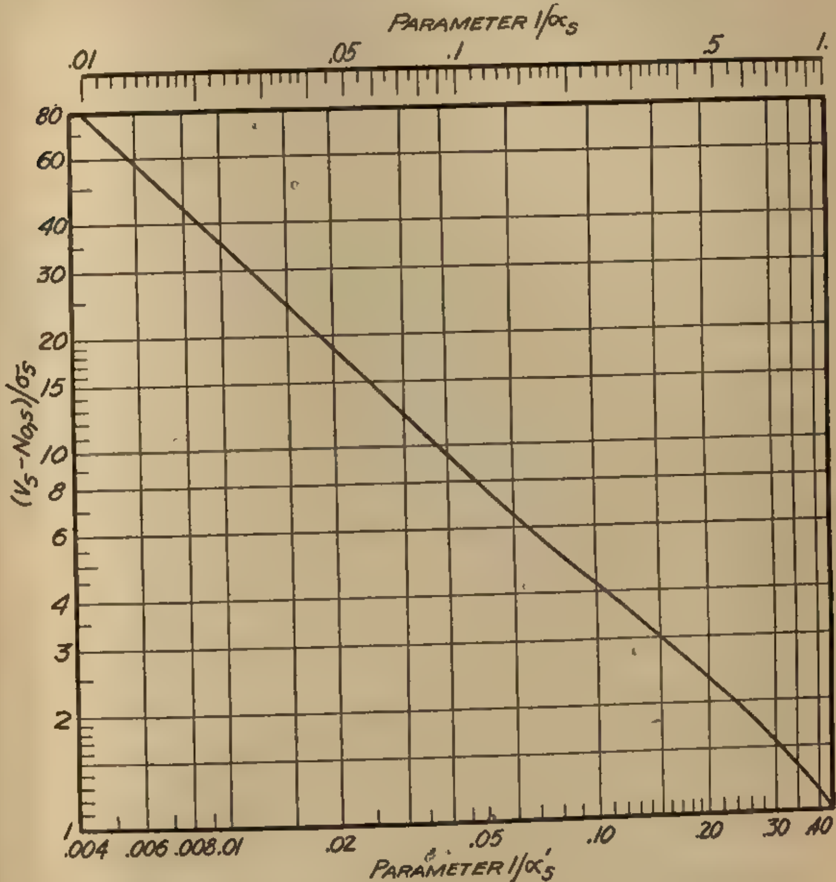


FIG. 6. Estimation of sensitivity limit $N_{0,S}$. (See Table 1.)

difference $\widehat{V}_S - \widehat{N}_{0,S}$ increases. For constant mean and standard deviation and increasing skewness the difference decreases.

The equations (3.6) and (3.7) lead to a new criterion which determines whether the minimum life $\widehat{N}_{0,S}$ vanishes or not, since

$$\widehat{N}_{0,S} = 0 \text{ if } \overline{N}_S + s_S(N)(A(\widehat{\alpha}_S) - B(\widehat{\alpha}_S) = 0.$$

This condition may be written from (3.6′) and (3.7′)

(4.7) $\widehat{N}_{0,S} = 0$; if $\overline{N_S^2}/\overline{N}_S^2 = \Gamma(1 + 2/\widehat{\alpha}_S)/\Gamma^2(1 + 1/\widehat{\alpha}_S)$.

The minimum life is taken to be zero, if the equality is fulfilled within the errors of random sampling. The same assumption must be made if the minimum life turns out to be negative.

Finally the limiting number of cycles $N_{\omega,S}$ for which the probability of survival is practically zero, is obtained from (2.9), (3.7) and (3.6) as

$$N_{\omega,S} = \overline{N}_S + \sigma_S[A(\alpha_S) + B(\alpha_S)(e^{2.5/\alpha_S} - 1)]. \tag{4.8}$$

It may be estimated by replacing the population values $\overline{N}_S$ and α_S by the sample values and using the functions $A(\alpha_S)$ and $B(\alpha_S)$ given in Table 1.

If the computed minimum life is longer than the smallest number of cycles at fracture, or if the largest observed number exceeds $N_{\omega,S}$, it must be taken into account in the evaluation of such contradictions that the estimations of $N_{0,S}$ and $N_{\omega,S}$ are subject to considerable errors of random sampling, which so far are unknown. This holds also for the different criteria to be used to prove or disprove the existence of a non-vanishing minimum life.

5. FATIGUE IN NICKEL AND ALUMINUM

Table 2 summarizes the observed number of cycles to fracture at

TABLE 2

NICKEL.—REVERSED TORSION WIRE TESTS (RAVILLY [5])

Fatigue Life N, Thousands of Cycles

Elastic Stresses S in kg/mm.^{2}

Specimen Number m	Plotting Position $1-\frac{m}{n+1}$	$S=\pm 15.5$	± 18.0	± 21.5	± 25.5	± 30	± 33	± 37.5	± 44.5	± 49	± 56
1	0.9524	1,040	424	208	118.0	84.3	58.4	40.0	23.0	17.5	11.0
2	0.9048	1,080	441	221	120.0	87.2	63.2	43.9	25.0	20.2	11.8
3	0.8571	1,110	442	223	140.0	87.2	64.4	45.4	26.7	20.5	12.6
4	0.8095	1,140	444	228	144.0	87.7	65.1	46.5	26.9	20.6	13.0
5	0.7619	1,150	449	228	144.5	95.3	68.0	47.1	28.4	20.7	13.1
6	0.7143	1,160	450	242	147.0	95.9	69.5	48.0	28.5	20.8	13.5
7	0.6667	1,170	456	244	148.0	96.5	69.7	48.2	28.9	20.8	13.6
8	0.6190	1,180	461	247	151.0	98.5	70.5	48.5	29.2	20.9	13.8
9	0.5714	1,180	465	248	153.0	101.4	71.0	49.3	29.8	21.0	14.0
10	0.5238	1,200	477	253	154.0	102.4	72.0	49.9	29.9	21.1	14.1
11	0.4762	1,210	480	253	157.0	104.0	73.3	51.2	29.9	21.5	14.2
12	0.4286	1,225	489	254	158.0	105.1	74.0	51.5	30.0	21.9	14.3
13	0.3810	1,240	492	255	159.0	105.9	75.1	52.3	30.3	22.1	14.4
14	0.3333	1,270	499	257	164.0	106.0	76.0	52.5	30.5	22.3	14.5
15	0.2857	1,280	503	262	165.0	106.0	76.0	53.7	31.0	22.3	14.6
16	0.2381	1,280	510	264	165.0	106.2	77.5	54.5	31.8	22.7	14.7
17	0.1905	1,310	510	287	170.0	106.6	78.0	54.8	32.0	22.9	14.8
18	0.1429	1,350	516	289	172.0	108.0	78.0	55.0	33.0	23.1	15.8
19	0.0952	1,460	534	292	173.0	108.0	79.2	55.7	33.6	23.7	16.1
20	0.0476	1,620	553	294	190.0	108.2	80.0	57.0	34.7	24.1	16.4

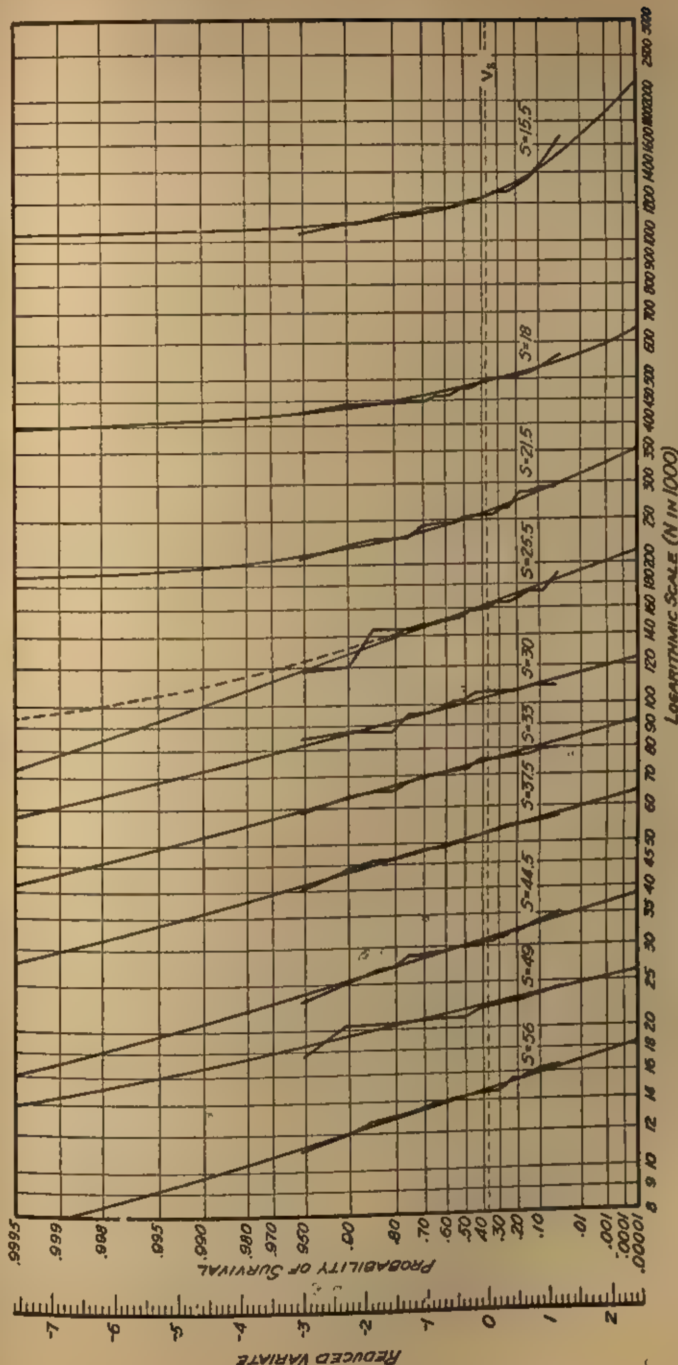


FIG. 7. Fatigue tests of nickel wire in reversed torsion. (See Tables 2 and 3.)

different stress amplitudes for nickel. These data, taken from Ravilly's observations [5] are traced on logarithmic extremal probability paper in Figure 7 using the plotting positions (1.7).

Table 3 gives the estimations of the three parameters. The estimates for the minimum life $N_{0,S}$ and the characteristic number V_S diminish with increasing stress level as might be expected. The estimated values of the scale parameters $1/\alpha_S'$ indicate that in most cases the mode ex-

TABLE 3

PARAMETERS FOR NICKEL WIRE

Non-Linear Theory

Stress Level (kg. per mm.²)	S	±15.5	±18.0	±21.5	±25.5
Number of Spec.	n	20	20	20	20
Mean	$\bar{N}_S$	1,232.75	479.75	252.45	154.625
Stand. Dev.	s_S	132.7198	34.7046	24.4228	17.0979
Third Moment	$m_{3,S}$	2,861,621.	13,422.2	2,524.58	−1,712.34
Scale Parameter	$1/\widehat{\alpha}_S'$	0.31438	0.16791	0.14554	0.07604
Char. Number	$\widehat{V}_S$	1,249.93	490.29	260.48	161.44
Sensitivity Limit	$\widehat{N}_{0,S}$	1,051.63	396.13	185.56	70.24

Linear Theory

Stress Level (kg. per mm.²)	S	±25.5	±30	±33	±37.5	±44.5	±49	±56
Number of Spec.	n	20	20	20	20	20	20	20
Mean Logarithm	$\overline{\log N}$	5.18664	4.99871	4.85573	4.69950	4.47015	4.33224	4.14468
Geom. Std. Dev.	$s(\log N)$	0.04850	0.03490	0.03594	0.03813	0.04118	0.02996	0.04112
Scale Parameter	$1/\widehat{\alpha}_S'$	0.04564	0.03284	0.03382	0.03588	0.03875	0.02819	0.03869
Log. Char. Number	$\log \widehat{V}_S$	5.21001	5.01558	4.87309	4.71793	4.49005	4.34672	4.16455
Char. Number	$\widehat{V}_S$	162.21	103.65	74.66	52.23	30.91	22.22	14.61

ceeds the median, a fact that contradicts the usual assumptions concerning the skewness of the distribution functions in fatigue. Within the linear theory, the estimates of $1/\alpha_S'$ do not show any systematic dependence upon S and the hypothesis that their variation is due to chance seems admissible on the basis of the considered tests.

Figure 7 shows that for nickel tested in reversed torsion the linear theory gives an excellent fit for stress levels equal to or exceeding 25.5 kg. mm.$^{-2}$, for which, therefore, no minimum life appears to exist. For stress levels below 25.5 kg. mm.$^{-2}$, however, the observations can be better fitted by the three-parameter survivorship functions which is bending upwards for decreasing numbers of cycles, indicating the existence of a minimum life. For the sake of comparison the theoretical survivorship functions for the linear, two-parameter theory (1.1) and for the general, three-parameter theory (2.1) are both shown for the stress level 25.5 kg. mm.$^{-2}$.

It is interesting to note in Table 3 that the two estimates for $V_{25.5}$ are practically equal. The observed and the two theoretical survivorship functions in this case are also shown in Figure 8 in the conventional way, where the number of cycles is traced as abscissa and the survivorship function as ordinate, both in linear scales. Again, no preference can be given to either theory.

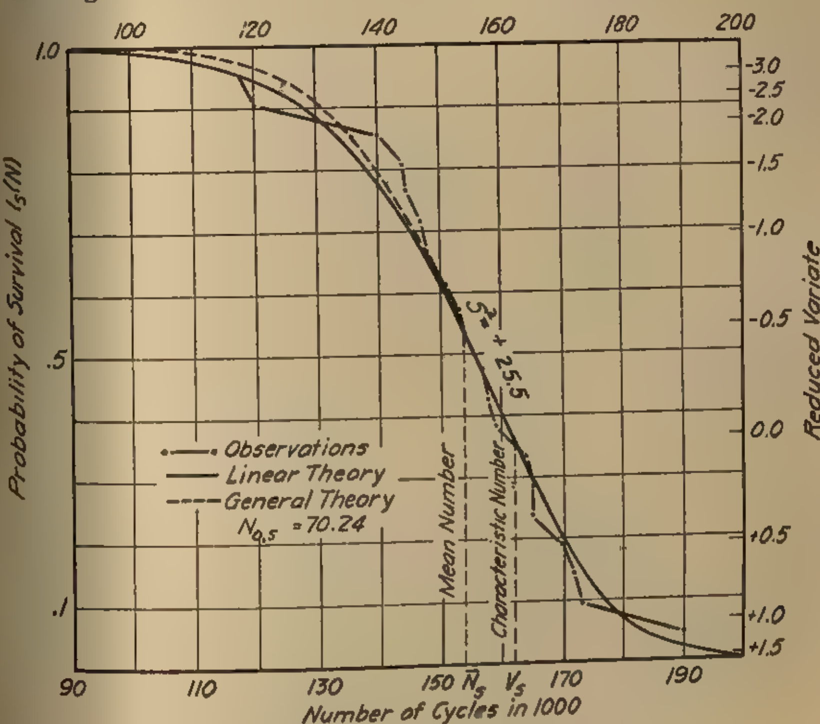


FIG. 8. Survivorship function for nickel at $S = \pm 25.5$ kg. mm^{-2}. (See Tables 2 and 3.)

Table 4 shows the calculation of the three parameters for Ravilly's observations on aluminum wire for the three lowest stress levels. Figure 9 presents the data given in [2] together with the theoretical curves obtained from Table 4. The survivorship function for the stress level 5.75 kg. mm.$^{-2}$ clearly shows how the minimum life converges toward zero with increasing level of applied stress; if the minimum life is sufficiently low and the stress level sufficiently high the general (three-parameter) theory hardly differs from the linear one, since with increasing stress levels the minimum life approaches zero so rapidly that no distinction between the two theories is possible.

TABLE 4

THE THREE PARAMETERS FOR ALUMINUM WIRE [2]

Stress Level (kg. per mm.^{2})	S	±5.25	±5.5	±5.75
Number of Spec.	n	20	20	20
Mean	$\overline{N}_S$	1,140.50	552.15	217.30
Stand. Dev.	s_S	179.867	177.1485	57.4228
Third Moment	$m_{3,S}$	2,672,050.	1,088,203.	−26,247.78
Scale Parameter	$1/\hat{\alpha}_S'$	0.18952	0.14888	0.10240
Char. Number	$\hat{V}_S$	1,190.41	609.76	238.79
Sensitivity Limit	$\hat{N}_{0,S}$	751.47	76.75	1.39

Tables 3 and 4 indicate that the $1/\hat{\alpha}_S$ decrease with increasing stress levels for those stresses where the minimum life does not vanish. While similar relations have been reported by some investigators, other observations show no such variation. The above relation cannot, therefore, be accepted as well established. Since all estimated values of $1/\alpha_S$ are considerably below unity, the existence of an exponential distribution of fatigue failures appears unlikely, for the reasons given above.

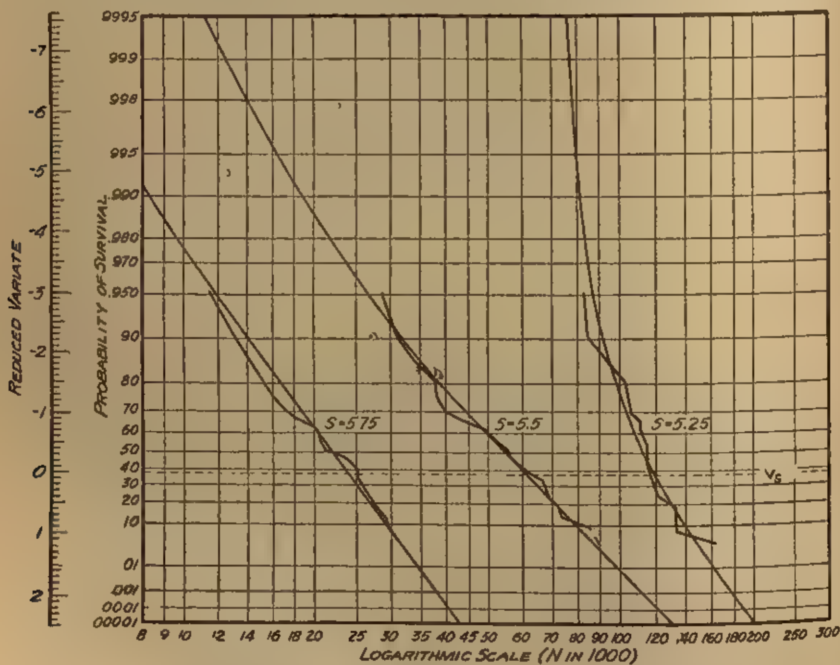


FIG. 9. Fatigue tests of annealed aluminum wire in reversed torsion. (See Table 4.)

CONCLUSIONS

The existence of an "incubation period" of fatigue, that is of a finite threshold number of cycles $N_{0,S}$ below which, at a given stress level, fatigue failure will not occur and at which the probability of survival is therefore *equal* to unity has been established for certain metals and stress amplitudes. This phenomenon has also been confirmed by fatigue studies on both hard and mild structural steel at stress levels near and below their static yield stress. Therefore a "sensitivity limit" in terms of cycles (minimum life) appears to be as real an aspect of fatigue as the sensitivity limit in terms of stress (endurance limit), the existence of which is quite generally recognized.

The present investigation is part of a research project on the basic aspects of fatigue, conducted at the Civil Engineering Research Laboratories of Columbia University in New York, and is supported by the Research Corporation, the Higgins Fund and, in part, by the Office of Ordnance Research.

REFERENCES

[1] Freudenthal, A. M., "Planning and Interpretations of Fatigue Tests," *American Society for Testing Materials, Special Technical Publication*. No. 121 (1953).

[2] Freudenthal, A. M., and Gumbel, E. J., "Statistical Interpretation of Fatigue Tests," *Proceedings of the Royal Society A.*, 216 (1953), 309–32.

[3] National Bureau of Standards, "Probability Tables for the Analysis of Extreme Value Data," *Applied Mathematics Series 22*, Washington (1953).

[4] Neyman, J. and Pearson, E. S., "On the Use and Interpretations of Certain Test Criteria," *Biometrika*, 20 (1928), 175–240.

[5] Ravilly, E., "Contributions à l'étude de la rupture des fils métalliques, soumis à des torsions alternées," *Publications Scientifiques et Techniques du Ministère de l'Air*. No. 120, Paris (1938), 52–70.

[6] Weibull, W., "The Phenomena of Rupture in Solids," *Handlingar Ingeniörs Vetenskaps Akademien No. 153*, Stockholm (1939).

[7] Weibull, W., "A Statistical Representation of Fatigue Failure in Solids," *Trans. Royal Institute of Technology*. No. 27, Stockholm (1949).

POINT ESTIMATES OF ORDINATES OF CONCAVE FUNCTIONS

CLIFFORD HILDRETH*
North Carolina State College

A method is developed for obtaining maximum likelihood estimates of points on a surface of unspecified algebraic form when ordinates of the points are required to satisfy a set of linear inequalities. A production function with one variable input is considered in some detail. In this case the restrictions follow from the assumption of non-increasing returns. An illustrative computation is worked out using a procedure based on equivalence between the estimation problem and a certain saddle point problem. Alternative procedures for production functions with two variable inputs are sketched.

1. INTRODUCTION

ECONOMISTS are frequently in the position of having fairly strong presumptions that relations among variables with which they deal satisfy certain qualitative restrictions, but they seldom have very good grounds for saying that a particular algebraic form is appropriate for representing a given relation. Diminishing marginal productivity of inputs in production relations, downward slope of demand relations and homogeneity of certain demand and production relations are examples of properties which economists often assume.

Unfortunately, statistical procedures available to economists typically require that they completely ignore many of their *a priori* presumptions in performing their statistical analyses or, alternatively, that they rather arbitrarily assume that a particular algebraic form satisfactorily represents the relation being investigated. In this paper procedures are suggested for estimating points on a production surface of unspecified algebraic form from data on outputs produced by various combinations of inputs when the inputs are subject to diminishing returns. An example involving a single variable input is worked out as an illustration. More generally one could apply the approach presented here to a variety of situations in which an investigator knows some properties of a relation being studied but does not have sufficient information to put the relation into any simple parametric form. Such situations are not uncommon so applications of the general approach may arise in several fields. However, the author's closer familiarity with

* The author is indebted to staff members of the Cowles Commission and to L. J. Savage of the Committee on Statistics, University of Chicago, for suggestions and critical comments. This paper will be reprinted as Journal Article No. 551 of the North Carolina Agricultural Experiment Station.

problems from economics makes it convenient to refer to this field when a more specific context is wanted.

The procedures to be outlined should be regarded as supplementing rather than supplanting existing techniques. When an investigator is reasonably sure that a particular parametric form will satisfactorily represent the relevant properties of his relation, there will ordinarily be advantages of both efficiency and convenience in using the form. Methods have also been developed for investigating particular aspects of a relation with only mild assumptions as to its form. Davis [4, Ch. 6] gives examples of smoothing devices that depend on local properties of a function. Robbins and Monro [13] have presented a stochastic method for approximating the value of a control variable associated with a selected mean response.

Procedures for estimating the locus of an extreme value of a function and for approximating its properties in that neighborhood have been suggested by Hotelling [7] and extended and refined by Friedman and Savage [5] and Box and Wilson [2].[1] These seem particularly useful if the investigator is interested in a fairly small region about the extreme value and if he has either a pretty good *a priori* notion of where to find the extremum or has the opportunity to draw a fairly large sequential sample. If some estimates of value of the function over a rather large region are needed, if fixed samples that cannot easily be repeated are the main source of information, and/or if there exists appropriate *a priori* information to be taken into account, then procedures like those suggested in what follows may have advantages. There is nothing to prevent an investigator from combining certain features of the present analysis with suggestions of Hotelling and the others if some opportunity exists to analyse certain initial data and then plan for the collection of additional observations.

The reader will have a better basis for judging possible applications after the illustrative production analysis has been presented. Statistical problems arising from the generation of variables by simultaneous economic relations are not considered in the present discussion nor are types of statistical inference other than point estimation. However, conventional tests of hypotheses could be used in many of the contemplated situations.

2. PRODUCTION FUNCTIONS WITH ONE VARIABLE INPUT

Consider a production relation of the form

$$y = \phi(z) + u \tag{2.1}$$

[1] A brief review of these procedures has recently been published by Anderson [1].

where y represents output, z represents variable input and u is a random disturbance. In general z may be a vector with as many components as there are types of variable input but for the present we assume that all inputs except one are held constant, or as nearly constant as physical conditions permit. If inputs are defined broadly enough to include all factors influencing output, then variations in u may be attributed to unavoidable and unobserved variations in some of the constant inputs. It is assumed that variations in u approximate independent drawings from a normal distribution with zero mean and finite variance.

An investigator is considered to have observations on y and z for N selected values of z. Let the values of z be arranged in increasing order and denoted by $z_1, z_2, \cdots, z_n, \cdots, z_N$. For each level of input there may be several trials and corresponding observations of output. Let T_n be the number of trials at level of input z_n and let y_{nt} be the observed output for the tth trial at this level. We have

$$y_{nt} = \phi(z_n) + u_{nt} \qquad \begin{array}{l} n = 1, 2, \cdots, N \\ t = 1, 2, \cdots, T_n. \end{array} \tag{2.2}$$

An economist with such data is principally interested in the possibility of drawing inferences about which levels of input are most profitable for various combinations of prices of output and of variable input (or conditions of demand for output and supply of input). Frequently such inferences have been drawn by assuming that the function $\phi(z)$ can be approximated by some given algebraic form with several unknown parameters to be estimated from the data. Estimates of the parameters are inserted into the form to obtain an estimated relation and this estimated relation is then used to calculate most profitable levels of input for chosen combinations of prices.

The chief difficulty with this procedure is that the inferences often depend critically upon the algebraic form chosen. It is not uncommon to find that alternative forms fit the data almost equally well but have very different implications for the most profitable level of input. One example of this may be found in a study by Paul R. Johnson [8] which will be utilized further in Section 3. A recent article by Prais [12] emphasizes the critical importance of the form of equation chosen to represent a demand relation. This problem would arise more frequently in economic literature if it were given attention commensurate with its importance. Many economists uncritically accept the appropriateness of a functional form chosen largely on the basis of convention or convenience; others try several forms for their relations but report only the one that in some sense "looks best" *a posteriori*.

If the investigator knew the prices of y and z (or more generally if he knew the relevant demand relation for y and supply relation for z) and if these were expected to remain fixed during the period that his statistical results were to be used, then he could express net revenue as a function of z and regard this as the relation to be studied. If, in addition, he could proceed to draw new observations of y, and therefore of net revenue, for chosen values of z, then he could apply the Hotelling (Friedman-Savage or Box-Wilson techniques could be used if z were a vector) procedure for estimating the point of maximum net revenue. If he wished the analysis to apply to various price (or demand and supply) situations or if he needed to draw some inferences before obtaining new observations, then the above techniques would be difficult to apply.

To compare all levels of input the researcher would, of course, have to make a very specific assumption about the form of $\phi(z)$. However, in many studies this is not really necessary. If a reasonable basis can be found for comparing the profitabilities of levels of input for which data exist, such comparisons will often determine the optimal level of input to as close an approximation as the available data permit. The results are typically intended for use as a guide to actual producers. Conditions faced by these producers can never be exactly duplicated in experiments so there is always some error in transferring experimental results to commercial situations. If the data are gathered by surveying actual producing units, fundamentally the same problem will exist. There will always be some more or less relevant discrepancies between conditions faced by sample producers at the times they are surveyed and conditions faced by producers who ultimately use the results at the time their applications are made. Thus complete accuracy in the determination of optimal input for the conditions represented by the data would always be superfluous even if it were possible. Furthermore, if comparisons among the observed levels of input are too crude when these considerations are taken into account, it is ordinarily possible to supplement the data and to obtain observations for additional levels in what is expected to be the relevant region. Indeed, it may frequently happen that an indication of the kinds of new observations that will prove useful may be one of the most valuable results of an initial examination of production data.

Let η_n be the expected value of output at input level z_n.

$$\eta_n = \phi(z_n) \qquad n = 1, 2, \cdots, N. \tag{2.3}$$

We seek to construct a reasonable procedure for obtaining estimates of the η_n; these can then be translated into estimates of expected profita-

bility of the z_n for any given price combinations. Our estimates will be derived by the method of maximum likelihood. However the same results could be obtained by least squares and possibly other methods. Of course, if there were no *a priori* restrictions on ϕ, then the maximum likelihood estimates of the *ordinates* η_n would just be the means of observed outputs for the appropriate levels of input, i.e.,

$$\tilde{\eta}_n = \bar{y}_n = \frac{1}{T_n} \sum_{t=1}^{T_n} y_{nt} \tag{2.4}$$

where the $\tilde{\eta}_n$ might be called limited information maximum likelihood estimates.[2] To obtain full maximum likelihood estimates (here denoted by $\hat{\eta}_n$) the investigator must maximize the likelihood function subject to all of the *a priori* restrictions he feels justified in imposing.

As was indicated earlier, in many production situations the investigator will feel that inputs are subject to decreasing returns. This is equivalent to assuming that ϕ is concave and yields the following restrictions on the ordinate[3]—

$$\frac{\eta_{n+1} - \eta_n}{z_{n+1} - z_n} \geq \frac{\eta_{n+2} - \eta_{n+1}}{z_{n+2} - z_{n+1}} \qquad n = 1, 2, \cdots, N-2. \tag{2.5}$$

For such cases it is desired to maximize the likelihood function subject to (2.5). The logarithm of the likelihood function is given by

$$L(\eta, \sigma^2) = -T/2 \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{t=1}^{T_n} (y_{nt} - \eta_n)^2 \tag{2.6}$$

where $T = \sum_{n=1}^{N} T_n$ and $\eta = (\eta_1 \eta_2 \cdots \eta_n)$. Since σ^2 is not restricted its estimator can be obtained by differentiation, yielding[4]—

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{n=1}^{N} \sum_{t=1}^{T_n} (y_{nt} - \hat{\eta}_n)^2. \tag{2.7}$$

where the $\hat{\eta}_n$ are those values which maximize L and satisfy (2.5); thus they minimize the double sum on the right of (2.6) under the restrictions (2.5). We may write

---

[2] For an analogous use of this term in another context see Koopmans and Hood [9].

[3] The investigator may typically feel justified in assuming strict concavity in which case the strict inequalities would hold in (2.5). However, if these restrictions are imposed, the likelihood function may not have a maximum but only a least upper bound. Maximizing the likelihood function subject to the restrictions given is equivalent to finding the least upper bound in the region defined by corresponding strict inequalities.

[4] This estimator may be expected to have a substantial downward bias when the T_n are small. Its distribution has not yet been investigated.

$$\sum_{n=1}^{N}\sum_{t=1}^{T_n}(y_{nt}-\eta_n)^2=\sum_{n=1}^{N}\sum_{t=1}^{T_n}(y_{nt}-\bar{y}_n)^2+\sum_{n=1}^{N}T_n(\eta_n-\bar{y}_n)^2. \tag{2.8}$$

The terms of the first summation over n on the right do not depend on η and may be ignored. Let x be an N-dimensional vector with elements

$$x_n=\eta_n-\bar{y}_n \qquad n=1,2,\cdots,N. \tag{2.9}$$

The problem is then to minimize the weighted (by the T_n) sum of squares of the x_n subject to the requirements that the $\hat{\eta}_n$ satisfy (2.5). Equivalent restrictions on the x_n are given by

$$\begin{aligned}-\frac{1}{\Delta_{n+1}}x_n+\left(\frac{1}{\Delta_{n+2}}+\frac{1}{\Delta_{n+1}}\right)x_{n+1}-\frac{1}{\Delta_{n+2}}x_{n+2}-\frac{1}{\Delta_{n+1}}\bar{y}_n\\+\left(\frac{1}{\Delta_{n+2}}+\frac{1}{\Delta_{n+1}}\right)\bar{y}_{n+1}-\frac{1}{\Delta_{n+2}}\bar{y}_{n+2}\geqq 0\end{aligned} \tag{2.10}$$

where $\Delta_{n+2}=z_{n+2}-z_{n+1}$, $\Delta_{n+1}=z_{n+1}-z_n$, and $n=1,2,\cdots,N-2$.

It will simplify the discussion to state the problem in matrix notation

(2.11) Find an $\hat{x}\epsilon A$ such that $\hat{x}D\hat{x}'\leqq xDx'$ for all $x\epsilon A$ where[5]

(i) A is the set of all vectors x that satisfy $A\,x'+b'\geqq 0$ and

(ii) $$A=\begin{bmatrix}\frac{-1}{\Delta_2} & \left(\frac{1}{\Delta_3}+\frac{1}{\Delta_2}\right) & \frac{-1}{\Delta_3} & 0 & \cdots & 0 & 0 & 0\\ 0 & \frac{-1}{\Delta_3} & \left(\frac{1}{\Delta_4}+\frac{1}{\Delta_3}\right) & \frac{-1}{\Delta_4} & \cdots & 0 & 0 & 0\\ & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & \frac{-1}{\Delta_{N-1}} & \left(\frac{1}{\Delta_N}+\frac{1}{\Delta_{N-1}}\right) & \frac{-1}{\Delta_N}\end{bmatrix}$$

(iii) $b'=A\bar{y}'$

(iv) $$D=\begin{bmatrix}T_1 & & & \\ & T_2 & & 0\\ 0 & & \ddots & \\ & & & T_N\end{bmatrix}.$$

[5] In the special case in which the input levels are equally spaced and the same number of trials exist for each level we may take $D=I$ and

$$A=\begin{bmatrix}-1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0\\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0\\ & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1\end{bmatrix}$$

since multiplication of D or any row of A by a positive constant does not change the problem. In this case, multiplication of a column vector by A yields the negative of the second differences of the elements of the vector.

The investigator can readily obtain D, A, b from his input-output data. We seek a way to compute $\hat{x}$ and thereby $\hat{\eta}$. Iterative procedures have generally been found most useful for this kind of computation. One might, for example, adapt a gradient method of minimization[6] to this problem or one might choose an arbitrary x satisfying the restrictions and proceed to minimize the form xDx' with respect to each component of x in turn holding the other components constant and observing the relevant restrictions at each stage. It is argued in Section 4 that the latter process would, in fact, converge to the minimizing vector.

However, for the particular problem stated in (2.11) it is possible to develop a more economical approach by using the fact that problems of extremization subject to inequalities have commonly been found to be equivalent to saddle point problems similar to those encountered in the Theory of Games. By a Theorem of Kuhn and Tucker [11, pp. 487, 491–2] the minimization problem stated in (2.11) is equivalent to the following saddle point problem—

(2.12) Find vectors $\hat{x}$, $\hat{v}$ such that

$$\phi(\hat{x}, v) \leq \phi(\hat{x}, \hat{v}) \leq \phi(x, \hat{v})$$

for all x and for all $v \geq O$ where

$$\phi(x, v) = xDx' - v(Ax' + b'). \tag{2.13}$$

Some of the general methods that have been developed for minimax problems could doubtless be adapted to this case. However, the following procedure seems exceedingly simple and is used in the example of Section 3.

Since D is positive definite, $\phi(x, v)$ has, for any v, a unique minimum with respect to x which may be found by differentiation.

$$\frac{\partial \phi}{\partial x} = 2Dx' - A'v'. \tag{2.14}$$

Setting the derivatives equal to zero yields

$$x' = \tfrac{1}{2}D^{-1}A'v' \tag{2.15}$$

and substituting into (2.13),

$$\min_x \phi(x, v) = \phi^*(v) = -\tfrac{1}{4}vAD^{-1}A'v' - vb'. \tag{2.16}$$

To find the non-negative $N-2$ dimensional vector $\hat{v}$ that maximizes this expression it is convenient to consider an equivalent minimum

[6] See, for example, Chernoff and Divinsky [3, pp. 246–7].

problem. Let

$$C = AD^{-1}A' \quad \text{and} \tag{2.17}$$

$$\theta(v) = -2\phi^*(v) = \tfrac{1}{2}vCv' + 2vb'. \tag{2.18}$$

Clearly $\hat{v}$ minimizes $\theta(v)$. Since A has $N-2$ linearly independent rows, it may be noted that C is positive definite. The procedure to be followed in finding $\hat{v}$ is an iterative one. An initial value, say $v^{(0)} = (v_1^{(0)} v_2^{(0)} \cdots v_{N-2}^{(0)})$, is chosen as a starting point for the iteration. Holding all components of v except the first fixed at their values given by $v^{(0)}$, the non-negative value of v_1 which minimizes $\theta(v)$ is found. Call this value $v_1^{(1)}$. $\theta(v)$ is then minimized with respect to admissible values of the second coordinate holding v_1 fixed at $v_1^{(1)}$ and v_3 to v_{N-2} fixed at $v_3^{(0)}$ to $v_{N-2}^{(0)}$. This process of minimizing with respect to each coordinate in turn while holding others fixed at their last obtained values is continued until the desired degree of stability[7] is obtained.

Let v_k, $k=1, 2, \cdots, K$, by a given component of v where, of course, $K=N-2$. The procedure indicated above can be made more explicit by observing that the minimum of $\theta(v)$ with respect to any v_k is either attained where $v_k=0$ or where $\partial\theta/\partial v_k=0$. If the latter equation yields a non-negative value for v_k, then this is the minimizing value, otherwise $v_k=0$ is the minimizing value. We note

$$\frac{\partial\theta}{\partial v} = Cv' + 2b'. \tag{2.19}$$

At the pth stage of the iteration, we define $w_k^{(p)}$ as the value of the kth coordinate that would be obtained by setting $\partial\theta/\partial v_k=0$, i.e.,

$$w_k^{(p)} = -\sum_{i=1}^{k-1} \frac{c_{ki}}{c_{kk}} v_i^{(p)} - \sum_{i=k+1}^{K} \frac{c_{ki}}{c_{kk}} v_i^{(p-1)} - 2\frac{b_k}{c_{kk}} \tag{2.20}$$

where the c_{ki} are elements of C and $k=1, 2, \cdots, K$.

The value of the kth coordinate of v at the pth stage of the iteration is then obtained by taking

$$v_k^{(p)} = \max(w_k^{(p)}, 0). \tag{2.21}$$

For the production problem described earlier it is convenient to start the process by setting $v^{(0)}=0$. The process then generates a sequence of vectors which we denote by $\{v^{(m)}\}$, i.e.,

[7] It is really a certain level of accuracy that is desired. In many iterative processes one intuitively associates this with the observed degree of stability. It would be worthwhile however to investigate circumstances under which the two may differ.

$$
\begin{aligned}
v^{(1)} &= (v_1^{(1)} \quad 0 \quad 0 \cdots 0)\\
v^{(2)} &= (v_1^{(1)} \quad v_2^{(1)} \quad 0 \cdots 0)\\
&\vdots\\
v^{(K)} &= (v_1^{(1)} \quad v_2^{(1)} \quad v_3^{(1)} \cdots v_K^{(1)})\\
v^{(K+1)} &= (v_1^{(2)} \quad v_2^{(1)} \quad v_3^{(1)} \cdots v_K^{(1)})\\
&\vdots\\
v^{(pK+k)} &= (v_1^{(p+1)} \cdots v_k^{(p+1)} \quad v_{k+1}^{(p)} \cdots v_K^{(p)})\\
&\vdots\\
&\text{etc.}
\end{aligned}
\tag{2.22}
$$

The function to be minimized, $\theta(v)$, then defines a corresponding sequence of scalars which we indicate by $\{\theta(v^{(m)})\}$. In Section 5 it is shown that the sequence $\{\theta(v^{(m)})\}$ converges to a unique minimum. It will be seen that the proof depends essentially on the existence of a unique minimum, the boundedness of $\{v^{(m)}\}$, and the continuity of $\theta(v)$ and its first derivatives. This proof is deferred because some readers may wish to see the process illustrated and discussed before becoming involved in the mathematical details of the proof. Once $\hat{v}$ has been obtained, $\hat{x}$ may be found from (2.15) and the estimates of the ordinates, the $\hat{\eta}_n$, follow from (2.9). The distribution of these estimates has not as yet been investigated. In Section 3, the computing procedure just described is applied to illustrative production data.

3. AN ILLUSTRATIVE COMPUTATION

The primary purpose of the illustration is to show how the computing procedure developed in the previous section can be applied. While data from actual experiments are used, I have not examined the original reports of these experiments and do not have any firm judgment about the appropriateness of combining data from these various experiments in the simple analysis proposed here. For this reason I do not try to discuss the economic implications of the data but merely use them to illustrate a computing procedure.

The data are taken from corn fertility experiments conducted at North Carolina State College. These have been summarized in a bulletin by Krantz [10]. Corn yields that resulted from various applications of nitrogen fertilizer are available. Paul R. Johnson [8] has used the results for fitting production functions under several alternative assumptions about the algebraic form of the function. Prior to his analysis Johnson made an attempt to select experiments that would

provide observations of yields under fairly similar conditions in all respects except level of nitrogen applied. Only results from plots with closely related soil types and from years of "good" weather were used.[8] Johnson's data consisted partly of direct observations and partly of interpolated values. Only the direct observations are used in the present calculation.

TABLE I

FITTED EQUATIONS AND OPTIMAL INPUTS FROM JOHNSON STUDY

Fitted Equation	Nitrogen Application to Maximize Profit
$y=4.504\ (z+20)^{0.59}$	5340
$y=25.16+.7595\ z-.00209\ z^2$	164
$y=108-82.48\ (.9897)^z$	230

y represents yield in bushels, z represents nitrogen in lbs.

In addition to approximating the underlying functional relation, Johnson was interested in the optimal application of nitrogen when nitrogen costs $0.137 per lb. and corn sells for $1.75 per bushel. While all of his equations fitted the data reasonably well, they differed sub-

[8] Good weather is attributed by Johnson to years in which "the rainfall distribution was about normal and soil moisture conditions were not low enough to cause the leaves to roll during this period (a five-week critical period including the time of tasseling)." In principle one should take account of the weather effect in the statistical specification if it is believed to have significantly affected the observed yields. If this were done and the weather and input effects were assumed to be additive, (2.2) would be replaced by

$$y_{nmt} = \eta_n + \delta_m + u_{nmt} \tag{2.2'}$$

where $m=1, 2, \cdots, M$ is an index of the year of a particular observation and δ_m is the weather effect in that year. Let T_{nm} be the number of observations at input level z_n in year m. If the T_{nm} are equal for all n and m, then the weather effect causes no significant complication. If we impose the natural requirement that $\sum_m \hat{\delta}_m = 0$, we have

$$\text{(i)} \qquad \hat{\delta}_m = \frac{\sum_n \sum_t y_{nmt}}{\sum_n T_{nm}} - \frac{\sum_n \sum_m \sum_t y_{nmt}}{\sum_n \sum_m T_{nm}}$$

from $m=1, 2, \cdots, M$. The η_n can be obtained by minimizing

$$\text{(ii)} \qquad s = \sum_n \left[\eta_n - \frac{\sum_m \sum_t y_{nmt}}{\sum_m T_{nm}} \right]^2$$

subject to the restrictions in (2.5) and the procedure developed in Section 2 applies directly. If the T_{nm} are unequal, as in the present case, the situation is more complicated. It seemed undesirable to introduce these complications in the present illustration especially since an effort had previously been made to select homogeneous years.

stantially in their implications for the most profitable level of nitrogen. This is shown in Table I where the first column shows the relations with estimated parameters filled in and the second column indicates the corresponding level of nitrogen for highest profit.

Unless one has considerable *a priori* confidence in the appropriateness of a particular algebraic form, there is considerable uncertainty about the implication of the data for the decision as to level of input. To illustrate the alternative procedure developed in Section 2, the observations are first listed in Table II.

TABLE II

OBSERVED YIELDS AT SPECIFIED LEVELS OF NITROGEN

Nitrogen (lbs./acre) (x_n)	0	20	40	60	80	120	160	180
	9.9	43.4	44.9	52.2	79.0	72.0	81.5	74.7
	31.3	27.3	40.2	66.0	68.6	74.1	72.9	110.3
	32.0	35.3	96.9	74.0	59.8	78.8	117.1	102.7
	24.2	42.2	52.1	64.3	81.7	107.0	102.3	120.9
Yields (Bu./Acre) (y_{ni})	18.8	35.7	85.1	77.3	107.1	102.5	114.3	103.9
	25.0	50.1	63.6	34.0	48.5	68.7	70.2	98.2
	2.8	56.0	77.3	58.5	94.6	78.1	83.9	70.7
	17.4	42.1	63.6	49.5	101.8	12.4	104.9	70.7
	14.6	42.1		62.2	94.6	74.0	113.9	
	25.8			50.1		69.8	104.9	
	13.3					85.0		
	33.0					80.8		
	19.8					73.0		
	25.0					100.5		
	63.9					115.5		
	24.4					92.1		
	22.8					83.9		
	51.7					96.3		
	11.6					96.3		
	14.4							
	20.8							
	19.1							
	7.2							
	16.6							
	55.2							
	3.7							
	15.2							
Sum	619.5	374.2	523.7	588.1	735.7	1560.8	965.9	752.1
No. of Obs. (T_n)	27	9	8	10	9	19	10	8
Mean ($\bar{y}_n$)	22.94	41.58	65.46	58.81	81.74	82.15	96.59	94.01

From these data, the following are readily computed

$$D^{-1} = \begin{pmatrix} \frac{1}{T_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{T_2} & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \frac{1}{T_N} \end{pmatrix}$$

$$(3.1) \quad = \begin{pmatrix} .0370 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .1111 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .1250 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .1000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .1111 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .0526 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .1000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .1250 \end{pmatrix}$$

$$(3.2) \quad A = \begin{pmatrix} -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 3 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 3 & -2 \end{pmatrix}.$$

The meanings of the restrictions are not changed if any row of A is multiplied by any positive constant. This sometimes makes it possible to choose convenient numbers for elements. The A above differs from that defined in (2.11) in that the above has been obtained by multiplying the first three rows by 20 and the last three rows by 40. Continuing

$$(3.3) \qquad b' = Ay' = \begin{pmatrix} -5.24 \\ 30.53 \\ -29.58 \\ 45.45 \\ -14.03 \\ 19.60 \end{pmatrix}$$

$$C = AD^{-1}A'$$

$$(3.4) \quad = \begin{bmatrix} .6064 & -.4722 & .1250 & 0 & 0 & 0 \\ -.4722 & .7111 & -.4500 & .2000 & 0 & 0 \\ .1250 & -.4500 & .6361 & -.7333 & .1111 & 0 \\ 0 & .2000 & -.7333 & 1.4525 & -.4385 & .0526 \\ 0 & 0 & .1111 & -.4385 & .4215 & -.4052 \\ 0 & 0 & 0 & .0526 & -.4052 & 1.4526 \end{bmatrix}.$$

From these, formulas for the $w_k^{(p)}$ are readily obtained.

$$(3.5) \quad \begin{aligned} w_1^{(p)} &= .7787v_2^{(p-1)} - .2061v_3^{(p-1)} + 17.28 \\ w_2^{(p)} &= .6640v_1^{(p)} + .6328v_3^{(p-1)} - .2812v_4^{(p-1)} - 85.86 \\ w_3^{(p)} &= -.1965v_1^{(p)} + .7074v_2^{(p)} + 1.1527v_4^{(p-1)} \\ &\quad - .1746v_5^{(p-1)} + 93.00 \\ w_4^{(p)} &= -.1377v_2^{(p)} + .5048v_3^{(p)} + .3019v_5^{(p-1)} \\ &\quad - .0362v_6^{(p-1)} - 62.58 \\ w_5^{(p)} &= -.2636v_3^{(p)} + 1.0403v_4^{(p)} + .9613v_6^{(p-1)} + 66.57 \\ w_6^{(p)} &= -.0362v_4^{(p)} + .2789v_5^{(p)} - 26.99. \end{aligned}$$

From (3.5) the values of $v_k^{(p)}$ shown in Table III resulted when $v^{(0)}$ was taken equal to the zero vector. It is recalled that

$$(2.21) \qquad v_k^{(p)} = \max(w_k^{(p)}, 0).$$

TABLE III

SUCCESSIVE VALUES OF $v_k^{(p)}$

k \ p	1	2	3	4	5
1	17.28	0	0	0	0
2	0	0	0	0	0
3	89.60	85.50	85.31	85.30	85.30
4	0	0	0	0	0
5	42.95	44.03	44.08	44.08	44.08
6	0	0	0	0	0

Applying (2.15) we have

$$\hat{x}' = \tfrac{1}{2}D^{-1}A'\hat{v}' = \begin{pmatrix} 0 \\ 0 \\ -5.33 \\ 8.53 \\ -7.19 \\ 2.32 \\ -2.20 \\ 0 \end{pmatrix} \tag{3.6}$$

and recalling (2.9), the maximum likelihood estimates of the ordinates are given by

$$\hat{\eta}' = \hat{x}' + \bar{y}' = \begin{pmatrix} 22.94 \\ 41.58 \\ 60.13 \\ 67.34 \\ 74.55 \\ 84.47 \\ 94.39 \\ 94.01 \end{pmatrix}. \tag{3.7}$$

These are shown together with the original observations and their means in Figure 1.

While a comparison of profitability of various applications is subject to the qualifications mentioned at the beginning of the section, it may be worthwhile to note that, at prices of $1.75 for corn and .137 for nitrogen, 160 lbs. would be better than the other levels according to our estimates. To get a good determination of optimal input for about this price ratio there should clearly be more observations in the 120–220 lbs. interval, they should be more closely spaced, and weather effects should be taken into account (see fn. 9). Since economists are usually interested in the optimal production practices corresponding to a number of possible price situations, a useful interpretation of the results could be obtained by determining for each observed level of input those price combinations at which a particular level of input is most profitable. Such a treatment has been illustrated for hypothetical production data by Hildreth and Reiter [6]. One consequence of estimating points on a surface, instead of estimating parameters in an equation assumed

to represent the surface, is that interpolation or extrapolation to parts of the surface for which no observations are available depends directly on judgment rather than on an initial assumption about the algebraic form. In many contexts this should probably be counted an advantage of form-free estimation. Experienced persons who may have useful notions of the approximate behavior of a surface may have difficulty visualizing all of the implications of a particular choice of algebraic

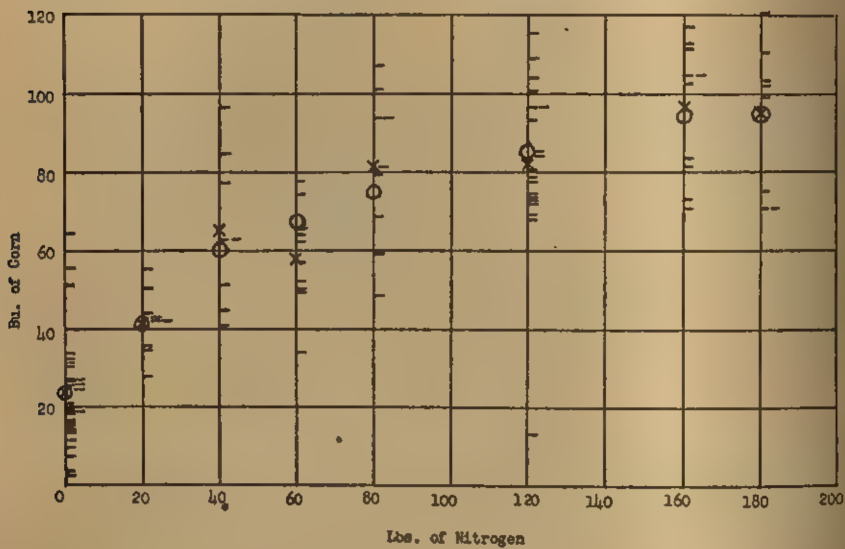


FIG. 1. Observations and Estimates.

— Observed yield (y_{ni})

× Mean yield ($\bar{y}_n$)

○ Maximum likelihood estimate of expected yield ($\tilde{\eta}_n$)

form. In addition, the form-free procedure allows necessary interpolations or extrapolations to be made at the stage of applying the results. At this stage various adjustments for possible discrepancies between experimental and commercial conditions need to be considered and advice of persons familiar with the circumstances of a particular application is likely to be available.

6. PRODUCTION FUNCTIONS WITH TWO VARIABLE INPUTS

Procedures similar to the one just illustrated could be applied to other problems of estimating an unknown coordinate of points on a

surface about which the investigator has qualitative *a priori* information, provided the *a priori* information could be translated into a set of linear inequalities (equalities would produce no complication and may be regarded as a special case) restricting the values of the unknown coordinate. It is likely that complications will sometimes arise in the translation of qualitative information into restrictions on the likelihood function. It may also be expected that different computing techniques will be required in some cases.

The proof of convergence of the iterative process used in Section 3 is seen in Section 5 to depend essentially on the existence of a unique minimum, the boundedness of $\{v^{(m)}\}$ and the continuity of $\theta(v)$ and its first derivatives. These conditions would all be met if the iteration were carried out on the vector x of the original problem (2.11) instead of the vector v of the dual problem. We could choose a vector $x^{(0)} = (x_1^{(0)}$ $x_2^{(0)} \cdots x_N^{(0)})$ to start the iteration, find an element $x_1^{(1)}$ which minimizes xDx' subject to $Ax'+b' \geqq 0$ with $x_2 \cdots x_N$ held fixed at their initial values, then minimize xDx' with respect to x_2, etc. This would have been more cumbersome (three x_n would appear in each restriction and, except at the ends, three restrictions would have to be examined each time an x_n were altered) than working with the dual problem in the case considered. However, in some problems it may be useful to consider an iteration on the original variables that enter the likelihood function.

For example, consider the case of a production function with two variable inputs, say

(4.1) $$y_{mnt} = \gamma(s_m, z_n) + u_{mnt} \quad \text{where}$$

s_m —the m^{th} level of one input
z_n —the n^{th} level of the other input
y_{mnt}—observed output of the t^{th} observation with input levels s_m and z_n
u_{mnt}—the value taken by the random disturbance on the t^{th} trial with input levels s_m and z_n,

and

$$m = 1, 2, \cdots, M; \; n = 1, 2, \quad , N; \; t = 1, 2, \cdots, T_{mn}.$$

To simplify the discussion we assume that the T_{mn} are all equal (the case of unequal numbers of trials could be covered by substituting $T_{mn} \cdot x_{mn}^2$ for x_{mn}^2 in (4.9) below the modifying subsequent equations accordingly). Let

$$\eta_{mn} = \phi(s_m, z_n) \quad \text{and} \tag{4.2}$$

$$x_{mn} = \eta_{mn} - \bar{y}_{mn} \quad \text{where} \tag{4.3}$$

$$\bar{y}_{mn} = \frac{1}{T_{mn}} \sum_{t=1}^{T_{mn}} y_{mnt}. \tag{4.4}$$

If we assume diminishing returns to each input, then the maximum likelihood estimates of the η_{mn} are obtained by minimizing the sum of squares of the x_{mn} subject to the restrictions imposed by diminishing returns. Let X denote the M rowed, N columned matrix with typical element x_{mn} and let $\bar{Y}$ be the $M \times N$ matrix with typical element $\bar{y}_{mn}$. The restrictions expressing diminishing returns to the first input are given by

$$A(X + \bar{Y}) \geqq 0 \quad \text{where} \tag{4.5}$$

A is of order $(M-2) \times M$ and is analogous to the A of Section 2; its elements are given by

$$\begin{aligned} a_{ij} &= 0 && \text{for } i > j \text{ and for } i < j - 2 \\ a_{ij} &= -\frac{1}{s_{i+1} - s_i} && \text{for } i = j \\ a_{ij} &= \frac{1}{s_{i+1} - s_i} + \frac{1}{s_{i+2} - s_{i+1}} && \text{for } i = j - 1 \\ a_{ij} &= -\frac{1}{s_{i+2} - s_{i+1}} && \text{for } i = j - 2. \end{aligned} \tag{4.6}$$

Similarly, the assumption of diminishing returns to the second input leads to the restrictions

$$B(X' + \bar{Y}') \geqq 0 \quad \text{where} \tag{4.7}$$

B is of order $(N-2) \times N$ with elements

$$\begin{aligned} b_{ij} &= 0 && \text{for } i > j \text{ and for } i < j - 2 \\ b_{ij} &= -\frac{1}{z_{i+1} - z_i} && \text{for } i = j \\ b_{ij} &= \frac{1}{z_{i+1} - z_i} + \frac{1}{z_{i+2} - z_{i+1}} && \text{for } i = j - 1 \\ b_{ij} &= -\frac{1}{z_{i+2} - z_{i+1}} && \text{for } i = j - 2. \end{aligned} \tag{4.8}$$

The problem is to find a matrix X such that the sum of squares of its elements is a minimum subject to (4.5) and (4.7). In the notation already introduced we wish to minimize $\gamma(X)$ subject to the diminishing returns restrictions letting

$$\gamma(X) = \sum_{m=1}^{M} \sum_{n=1}^{N} x_{mn}^2 = \operatorname{tr} XX' \tag{4.9}$$

where tr is an abbreviation for trace.

This could be done by iterating on the elements of X subject to the restrictions (4.5) and (4.7).

Alternatively one could consider the dual problem of finding the minimax of

$$\psi(X, V, W) = \operatorname{tr} XX' - \operatorname{tr} VA(X + \bar{Y}) - \operatorname{tr} WB(X' + \bar{Y}') \tag{4.10}$$

where V is an $N \times (M-2)$ matrix of coefficients to be determined and W is a similar matrix of order $M \times (N-2)$. We seek the minimum with respect to X and the maximum with respect to V and W subject to $V \geqq 0$, $W \geqq 0$.

As in the single input case, the minimizing values for X can be found by differentiation.

$$\frac{\partial \psi}{\partial X} = 2X' - VA - B'W' = 0 \tag{4.11}$$

$$X = \tfrac{1}{2}(A'V' + WB). \tag{4.12}$$

Substituting this into (4.10) yields

$$\begin{aligned}\psi^*(V, W) = &- \tfrac{1}{4} \operatorname{tr} VAA'V' - \tfrac{1}{2} \operatorname{tr} VAWB - \tfrac{1}{4} \operatorname{tr} WBB'W' \\ &- \operatorname{tr} VA\bar{Y} - \operatorname{tr} WB\bar{Y}'.\end{aligned} \tag{4.13}$$

To complete the analogy with the single input case we define

$$\begin{aligned}\theta(V, W) = - 2\psi^* = &\tfrac{1}{2} \operatorname{tr} (VAA'V' + 2VAWB + B'W'WB) \\ &+ 2 \operatorname{tr} VA\bar{Y} + 2 \operatorname{tr} WB\bar{Y}'\end{aligned} \tag{4.14}$$

which is to be minimized subject to the nonnegativity of the elements of V and W. One could proceed to iterate for the minimizing elements of W and V. However, if N and M are very large this will be a long process and might be no easier than to compute the original problem of minimizing $\gamma(X)$. It should also be noted that the proof of convergence in Section 5 does not apply to this case because the quadratic part of (4.13) is positive semi definite rather than positive definite. While it

seems a reasonable conjecture that the computation would converge, this problem needs further investigation. In any problem in which the number of linear inequalities exceeds the number of variables in the likelihood function, the quadratic part of the expression to be minimaxed in the dual problem will be positive semi definite and the convergence of the suggested iteration will have to be shown.

These and other complications make it hard to foresee which kinds of computing arrangements are likely to be most generally useful. As experience indicates more exactly the kinds of relations and restrictions to which applied workers want to apply methods like those developed here, it will be useful to give more attention to this problem.

5. PROOF OF CONVERGENCE

In this section we wish to show that the procedure suggested in Section 2 leads in the limit to a unique minimum of the function $\theta(v)$ for v in the closed positive orthant and therefore to estimates of ordinates which maximize the likelihood function in its restricted domain.[9] We recall that $\{v^{(m)}\}$ is a sequence of vectors in the positive orthant of a K dimensional Euclidean space, and that $v^{(m+1)}$ is obtained from $v^{(m)}$ by adjusting one element of v so as to minimize $\theta(v)$ subject to the conditions that the adjusted element remain non-negative and that other elements retain the values assumed in $v^{(m)}$.

Recalling also that

$$\theta(v) = \tfrac{1}{2}vCv' - 2vb' \tag{2.18}$$

where C is positive definite, we note that the sequence $\{\theta(v^{(m)})\}$ is non-increasing and bounded below, therefore it converges.

In addition, it can be argued that $\theta(v)$ has a unique minimum for v in the (closed) positive orthant. In the first place any minimizing point must lie in the intersection of the positive orthant and the ellipsoid $\theta(v) \leqq \theta(v^{(0)})$. Since this intersection is closed and bounded, a minimum is attained there. Suppose there were two minimizing points, say v^* and v^{**}. The line segment joining them would lie in the positive orthant and would also lie in the ellipsoid $\theta(v) \leqq \theta(v^*) = \theta(v^{**})$. Points in the interior of this ellipsoid correspond to lower values of $\theta(v)$ than points on the surface, i.e. $\theta(v) < \theta(v^*)$ for v in the interior of $[v:\theta(v) \leqq \theta(v^*).]$ Unless $v^* = v^{**}$, the line segment joining them contains interior points and the supposition that v^*, v^{**} were minimizing points is contradicted.

---

[9] The main features of this proof were suggested by Roy Radner.

We should next like to show that

$$\lim_{p\to\infty} \left| v_k^{(p)} - v_k^{(p+1)} \right| = 0 \qquad \text{for all } k. \tag{5.1}$$

Let $m=pK+k$. Then in passing from $v^{(m-1)}$ to $v^{(m)}$, only the kth coordinate of v changes. Consider the following—

$$\begin{aligned} \theta(v^{(m-1)}) - \theta(v^{(m)}) &= \tfrac{1}{2}c_{kk}(v_k^{(p)})^2 - v_k^{(p+1)})^2 \\ &+ (v_k^{(p)} - v_k^{(p+1)})\left[\sum_{i=1}^{k-1} c_{ki}v_i^{(p+1)} + \sum_{i=k+1}^{K} c_{ki}v_i^{(p)} + 2b_k\right]. \end{aligned} \tag{5.2}$$

Changing the superscript in (2.20) yields

$$w_k^{(p+1)} = -\sum_{i=1}^{k-1} \frac{c_{ki}}{c_{kk}} v_i^{(p+1)} - \sum_{i=k+1}^{K} \frac{c_{ki}}{c_{kk}} v_i^{(p)} + 2\frac{b_k}{c_{kk}}. \tag{2.20'}$$

Substituting into (5.2) we obtain

$$\begin{aligned} \theta(v^{(m-1)}) - \theta(v^{(m)}) &= \frac{c_{kk}}{2}(v_k^{(p)} - v_k^{(p+1)})(v_k^{(p)} + v_k^{(p+1)} - 2w_k^{(p+1)}) \\ &\geqq \frac{c_{kk}}{2}(v_k^{(p)} - v_k^{(p+1)})^2. \end{aligned} \tag{5.3}$$

To justify the mixed inequality we note that if $w_k^{(p+1)} \geqq 0$ then $v_k^{(p+1)} = w_k^{(p+1)}$ and the equality holds. If $w_k^{(p+1)} < 0$, then $v_k^{(p+1)} = 0$ and $(v_k^{(p)} - 2w_k^{(p+1)}) > v_k^{(p)} \geqq 0$ and the mixed inequality holds. (5.1) follows from (5.3) and the convergence of $\{\theta(v^{(m)})\}$.

Let $P_k(v)$ be the vector obtained from v by holding all but the k^{th} component fixed and minimizing $\theta(v)$ with respect to v_k. P_k is a continuous mapping of K dimensional Euclidean space into itself. In our sequence $\{v^{(m)}\}$ we have

$$v^{(m)} = P_{m\sim K}(v^{(m-1)}) \quad \text{where } m \sim K \text{ means } m \text{ modulo } K. \tag{5.4}$$

Let $\theta_{min} = \theta(v_{min})$ be the value of our function at its minimum for non-negative v. Let θ_∞ be the limit of our sequence $\{\theta(v^{(m)})\}$. We should like to show that $\theta_\infty = \theta_{min}$. The vector sequence $\{v^{(m)}\}$ is bounded. In particular the ellipsoid given by $\theta(v) = \theta(v^{(0)})$ contains the ellipsoids given by $\theta(v) = \theta(v^{(m)})$ and thus bounds the sequence. $\{v^{(m)}\}$ therefore has at least one limit point and contains a subsequence which converges to this limit. Let v^∞ be limit a of $\{v^{(m)}\}$ and let $\{v^{(r)}\}$ be a subsequence approaching v^∞. For each r identifying an element of the subsequence, let $m(r)$ identify the same element in the original sequence.

From the continuity of θ,

$$\theta_\infty = \theta(v^\infty). \tag{5.5}$$

We shall show that to suppose $\theta(v^\infty) \neq \theta_{\min}$ involves a contradiction.

If the supposition is true then it is possible to reduce $\theta(v)$ by changing a coordinate of v^∞, i.e.

$$\exists\, k \ni P_k(v^\infty) \neq v^\infty. \tag{5.6}$$

Let $\overline{K}$ be the set of all such k and let

$$\varepsilon = \min_{k \varepsilon \overline{K}} \left[\theta(v^\infty) - \theta(P_k(v^\infty))\right]. \tag{5.7}$$

Let $\overline{V}$ be the set of all v in the convex set bounded by the ellipsoid $\theta(v) = \theta(v^0)$. $\theta(v)$ is uniformly continuous over $\overline{V}$, i.e.,

$$\exists\, \delta > 0 \ni \|v - v^*\| < \delta \rightarrow |\theta(v) - \theta(v^*)| < \varepsilon \tag{5.8}$$

for all $v, v^* \,\varepsilon\, \overline{V}$.

We proceed to show that our original vector sequence, $\{v^{(m)}\}$, contains an element $P_k(v^{(m)})$ within δ of $P_k(v^\infty)$ for a $k \varepsilon \overline{K}$. It will then follow from (5.8) that $|\theta(P_k(v^{(m)})) - \theta(P_k(v^\infty))| < \varepsilon$. Since $\theta(P_k(v^\infty))$ is at least ε below θ_∞, $\theta(P_k(v^{(m)}))$ must also be less than θ_∞. But since θ_∞ is the limit of $\theta(v^{(m)})$, this is in contradiction to the definition of $\{v^{(m)}\}$.

From the continuity of the P_k we know

$$\exists\, \rho > 0 \ni \|v - v^\infty\| < \rho \rightarrow \|P_k(v) - P_k(v^\infty)\| < \delta \tag{5.9}$$

for all k.

From (5.1) we know that successive elements of $\{v^{(m)}\}$ can be made arbitrarily close together by making m sufficiently large. We also know that if r is sufficiently large, elements of the subsequence $\{v^{(r)}\}$ lie arbitrarily close to v^∞. Specifically we may say

$$\begin{aligned} &\exists\, M \ni m > M \rightarrow \|v^{(m+1)} - v^{(m)}\| < \frac{\rho}{K+1} \\ &\exists\, R \ni r > R \rightarrow \|v^{(r)} - v^\infty\| < \frac{\rho}{K+1}. \end{aligned} \tag{5.10}$$

Now consider an $\bar{r}$ such that $\bar{r} > R$, $\bar{m} = m(\bar{r}) > M$. The K elements of $\{v^{(\bar{m})}\}$ immediately following $v^{(\bar{m})}$ are all within ρ of v^∞. At least one of them is obtained from the preceding by applying P_k with $k \varepsilon \overline{K}$. Such an element, $P_k(v^{(\bar{m}+i)})$ with i an interger between 0 and $K-1$, lies within δ of $P_k(v^\infty)$ and reduces $\theta(v)$ below θ_∞, i.e.,

$$\left\|v^{(\bar{m}+i)} - v^{\infty}\right\| < \rho \qquad \text{for some } 0 \leqq i \leqq K - 1 \tag{5.11}$$

such that $v^{(\bar{m}+i+1)} = P_k(v^{(m+i)})$ for $k \varepsilon \underline{\bar{K}}$. From this and (5.9) we obtain

$$\left\|P_k(v^{(\bar{m}+i)}) - P_k(v^{\infty})\right\| < \delta \qquad \text{so that by (5.8)} \tag{5.12}$$

$$\left| \theta(P_k(v^{(\bar{m}+i)}) - \theta(P_k(v^{\infty}))\right| < \varepsilon. \qquad \text{Then from (5.7)} \tag{5.13}$$

$$\theta(P_k(v^{(\bar{m}+i)}) < \theta_{\infty}. \tag{5.14}$$

REFERENCES

[1] Anderson, R. L., "Recent advances in finding best operating conditions," *Journal of the American Statistical Association*, 48 (1953), 789–98.

[2] Box, G. E. P., and Wilson, K. B., "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society*, Series B, 13 (1951), 1–45.

[3] Chernoff, Herman and Divinsky, Nathan, "The computation of maximum-likelihood estimates of linear structural relations," Ch. X of *Studies in Econometric Method*, edited by Hood and Koopmans, New York, John Wiley and Sons, 1953.

[5] Friedman, Milton and Savage, L. J., "Planning experiments seeking maxima," Ch. 13 of *Techniques of Statistical Analysis*, edited by Eisenhart, Hastay and Wallis, New York, McGraw-Hill Book Co., 1947.

[6] Hildreth, Clifford, and Reiter, Stanley, "On the choice of a crop rotation plan," Ch. XI of *Activity Analysis of Production and Allocation*, edited by T. C. Koopmans, New York, John Wiley and Sons, 1951.

[7] Hotelling, Harold, "Experimental determination of the maximum of a function," *Annals of Mathematical Statistics*, 12 (1941), 20–45.

[8] Johnson, Paul R., "Alternative functions for analyzing a fertilizer-yield relationship," *Journal of Farm Economics*, 35 (1953), 519–29.

[9] Koopmans, Tjalling C. and Hood, William C., "The estimation of simultaneous linear economic relations," Ch. VI of *Studies in Econometric Method*, edited by Hood and Koopmans, New York, John Wiley and Sons, 1953.

[10] Krantz, B. A., *Fertilize Corn for Higher Yields*, Bulletin 366 of the North Carolina Agricultural Experiment Station, Raleigh, 1946.

[11] Kuhn, H. W. and Tucker, A. W., "Nonlinear programming," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, edited by Jerzy Neyman, Berkeley, University of California Press, 1951.

[12] Prais, S. J., "Non-linear estimation of the Engel Curves," *Review of Economic Studies*, 20 (1953), 87–104.

[13] Robbins, Herbert and Monro, Sutton, "A stochastic approximation method," *Annals of Mathematical Statistics*, 22 (1951), 400–7.

APPROXIMATE DISTRIBUTION OF THE RANGE IN THE NEIGHBORHOOD OF LOW PERCENTAGE POINTS*

MAURICE H. BELZ,† *University of Melbourne*

AND

ROBERT HOOKE, *Princeton University*

I. GENERAL

1. *Introduction*

THE investigations described below originated in a study of the upper-tail probabilities associated with the distribution of the range of a small sample from a normal population. If the cumulative distribution function for such a range is plotted on probability paper, as in Hald [3], it is observed that the curves (for different sample sizes) tend to become closely linear as the range increases. The same feature appears to characterize the distribution of the range for small samples from other continuous populations, for example, the χ^2 distributions for 2 and 4 degrees of freedom and the double negative exponential distribution; see Fig. 1. Attempts to discover the reason for this prop-

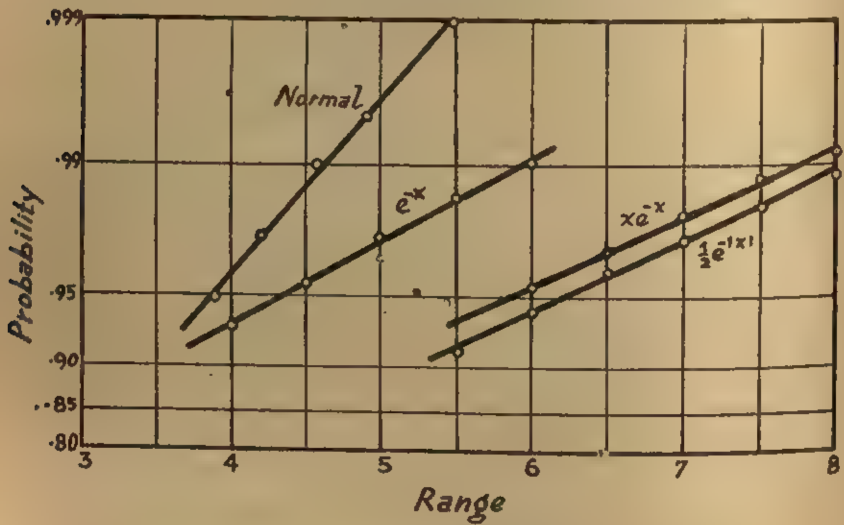


FIG. 1. Distribution of range of samples of 5 from various populations.

* Prepared in connection with research sponsored, in part, by the U. S. Office of Naval Research.

† Research Associate, Princeton University, 1952.

erty have led to some general approximation procedures for estimating upper-tail probabilities for the distribution of the range of small samples.

If $x_1 < x_2 < \cdots < x_n$ represent the ordered members of a sample of size n, the range is $x_n - x_1$. If this statistic is such that the probability of its being exceeded is low, say of the order of .05 or .01, the effect of x_n on x_1 might well be supposed to be small, whatever the value of n. This suggests, first, the investigation of an approximation based on the notion of complete independence of x_1 and x_n. Secondly, one may attach to this the assumption that x_1 and x_n are normally distributed, a result that is known to be approximately true, for small values of n, when the parent population is normal [6]. Results of the application of these procedures to small samples from various populations are reported below.

2. *Notation and Summary of Results*

The following notation is employed throughout. For $a>0$,

$P_1(a)$ = Probability that the range exceeds a;

$P_2(a)$ = approximation to the probability of the same event, based on the supposition that x_1 and x_n are independent, with means and variances characteristic of their actual distributions.

$P_3(a)$ = approximation to the probability of the same event, based now on the supposition that x_1 and x_n are independently and normally distributed, with means and variances characteristic of their actual distributions;

$R(a) = P_1(a)/P_2(a)$;

$R = \lim_{P_1(a) \to 0} R(a)$.

It is shown that, in general, R is not equal to 1, the value that might intuitively have been expected, so that the indicated approximations to $P_1(a)$ are $RP_2(a)$ and $RP_3(a)$. For any distribution defined over a finite interval, $R = (n-1)/n$. For distributions having just one infinite tail, R lies between $(n-1)/n$ and 1 and can, in fact, be as large as 1 or as small as $(n-1)/n$. These results appear also to be true in the case of distributions with two infinite tails, though no proofs have yet been obtained.

Numerical applications of the approximation procedures have been made to particular distributions, mostly with $n=5$ and with emphasis on those values of a which place $P_1(a)$ between .05 and .01, the interval of most interest in practice. The results are summarized in Table 1.

TABLE I

TAIL PROBABILITIES FOR THE DISTRIBUTION OF THE RANGE
($n=5$ unless otherwise noted)

Distribution	a	Probabilities: True $P_1(a)$	Approximate $P_2(a)$	Approximate $P_3(a)$	$R(a)$	R	$RP_2(a)$	$RP_3(a)$
Rectangular	.92	.0544	.0646	.1017	.842	.8	.0517	.0814
$f(x)=1$ for $0\leq x\leq 1$	.954	.0193	.0234	.0746	.825	.8	.0187	.0597
Symmetrical Triangular	1.485	.0490	.0583	.0643	.855	.8	.0467	.0514
$f(x)=\begin{cases}x \text{ for } 0\leq x\leq 1\\ 2-x \text{ for } 1\leq x\leq 2\end{cases}$	1.5	.0447	.0532	.0590	.840	.8	.0426	.0472
	1.8	.0013	.0016	.0077	.807	.8	.0013	.0062
Beta, parameters 2, 2	.7	.1403		.1565		.8		.1252
$f(x)=6x(1-x)$	.78	.0497		.0720		.8		.0576
for $0\leq x\leq 1$	.8	.0355		.0577		.8		.0461
	.85	.0124		.0317		.8		.0254
Chi-square, 2 d.f.	4.36	.0501	.0521	.0317	.962	.96	.0500	.0304
$f(x)=e^{-x}$ for $0\leq x<\infty$	5.06	.0251	.0262	.0076	.961	.96	.0251	.0073
Chi-square, 4 d.f. $f(x)=xe^{-x}$ for $0\leq x<\infty$	6	.0461	.0492	.0318	.937	.9262	.0456	.0295
Chi-square, $2m+2$ d.f. $f(x)=x^m e^{-x}/m!$ for $0\leq x<\infty$ (n unrestricted)						$\frac{n-1}{n}$ when $m\to\infty$		
Hyberbolic	50	.0568	.0571		.993	1	.0571	
$f(x)=1/(1+x)^2$ for $0\leq x<\infty$ ($n=3$)	100	.0292	.0293		.996	1	.0293	
Hyperbolic	6	.0544	.0570		.972	1	.0570	
$f(x)=2/(1+x)^3$ for $0\leq x<\infty$ ($n=3$)	10	.0233	.0238		.980	1	.0238	
Hyperbolic	2	.0916	.0976	.1998	.939	1	.0976	.1998
$f(x)=3/(1+x)^4$	2.7	.0508	.0530	.0840	.958	1	.0530	.0840
for $0\leq x<\infty$ ($n=3$)	3	.0407	.0424	.0542	.960	1	.0424	.0542
	4	.0214	.0222	.0089	.965	1	.0222	.0089
Hyperbolic $f(x)=(m-1)/(1+x)^m$, $m\geq 2$, for $0\leq x<\infty$ (n unrestricted)						1		
Double Negative Exponential	6	.0589	.0658	.0500	.894	.8	.0527	.0400
$f(x)=\frac{1}{2}e^{-\lvert x\rvert}$ for $-\infty<x<\infty$	8.2	.0095	.0110	.0011	.864	.8	.0088	.0009
Double Hyperbolic $f(x)=\frac{1}{2}(m-1)/(1+\lvert x\rvert)^m$, $m\geq 2$, for $-\infty<x<\infty$ (n unrestricted)						1		
Standard Normal	3.86	.0498	.0580	.0525	.858	(?)* .8	(?)* .0464	(?)* .0420
$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ ($n=5$)	4.00	.0377	.0440	.0384	.857	.8	.0352	.0307
for $-\infty<x<\infty$	4.20	.0248	.0294	.0238	.845	.8	.0235	.0190
	4.60	.0101	.0121	.0081	.835	.8	.0097	.0065
($n=8$)	4.29	.0497	.0543	.0474	.917	.875	.0475	.0415
	4.61	.0247	.0273	.0206	.905	.875	.0239	.0180
	[illegible]	.0099	.0111	.0065	.890	.875	.0097	.0057
($n=10$)	4.47	.0505	.0542	.0467	.930	.9	.0488	.0420
	4.79	.0247	.0267	.0195	.925	.9	.0240	.0176
	5.16	.0099	.0106	.0060	.915	.9	.0097	.0054
Poisson, mean 1	5	.0157	.0166		.946			
$p(x)=e^{-1}/x!$	6	.00256	.00271		.945			

* Question marks indicate that, although R is tak'n to b 0.8, this has not been established in this case.

3. *Conclusions*

Table I shows that $RP_3(a)$ provides a fair to good approximation to $P_1(a)$ if $P_1(a)$ is near .05 or if the parent distribution is normal. When the parent population is unknown it therefore seems reasonable to suppose that $RP_3(a)$ will be a good approximation to $P_1(a)$ provided that the distribution of the parent is not too far from normal. Even when this condition is not fulfilled, the variety of examples in Table I indicates that the approximation is likely to be good if $P_1(a)$ is near to .05.

The results of examples involving two-tailed distributions are consistent with the suggestion that R lies between $(n-1)/n$ and 1. In all the cases examined the ratio $R(a)$ was found to lie between this same pair of values. Failure to know the exact value of R is not very serious, for even when n is as small as 5 the range of R is only from .8 to 1, and whatever the value one takes for R in this range no error of any consequence is likely to be committed.

It is shown in Section 7 that when the parent population is unknown and cannot be assumed to be normal, the use of the approximation $RP_3(a)$ is likely to be considerably better than that obtained by the direct method of examining observed ranges, when the amount of data is small. This is supported by some results obtained from drawings from various populations.

The closeness of the approximations obtained for distributions defined over a finite interval as well as for the normal distribution suggests that the procedure might be usefully extended to contrasts other than the range. Tables of percentage points to cover quite simple contrasts are not now available, even for a normal parent, and some approximation procedure must accordingly be adopted. An investigation of the contrast $x_n-\frac{1}{2}(x_1+x_2)$ has been carried out along the above lines for samples of 5 from four different populations, and in Section 8 it is shown that the approximation $RP_3(a)$ is remarkably close to the true probability $P_1(a)$ when this is of the order .05 to .01. The analytical results are again supported by means of random drawings.

II. THE SUPPOSITION OF INDEPENDENCE—THEORETICAL RESULTS

4. *Expressions for $F_1(a)$, $P_2(a)$, $R(a)$*

Let $f(x)$, $F(x)$ be the density function and distribution function, respectively, and let (α, β) be the interval over which x is defined. Then we have the following results where, for simplicity, u is written in place of x_1 and v in place of x_n.

$$(1)\quad P_1(a) = \Pr(v-u>a), \qquad (a>0)$$

$$= \int_\omega n(n-1)f(u)\{F(v) - F(u)\}^{n-2}f(v)dudv,$$

$$(2) \quad P_2(a) = \int_\omega n^2 f(u)\{1 - F(u)\}^{n-1}F^{n-1}(v)f(v)dudv,$$

$$(3) \quad R(a) = \frac{P_1(a)}{P_2(a)} = \frac{n-1}{n} \cdot \frac{\int_\omega \{F(v) - F(u)\}^{n-2}f(u)f(v)dudv}{\int_\omega \{1 - F(u)\}^{n-1}F^{n-1}(v)f(u)f(v)dudv},$$

where ω is the relevant portion of the half-plane for which $v > u + a$.

In evaluating the integrals it is convenient to integrate first with respect to v, a distribution-free procedure, the terminals for the integral being $u+a$ and β. The integration with respect to u is then from α to $\beta - a$. We find the following expressions,

$$(4) \quad \begin{aligned} P_1(a) &= 1 - \{1 - F(\beta - a)\}^n \\ &\quad - n\int_\alpha^{\beta - a} f(u)\{F(u + a) - F(u)\}^{n-1}du, \end{aligned}$$

$$(5) \quad \begin{aligned} P_2(a) &= 1 - \{1 - F(\beta - a)\}^n \\ &\quad - n\int_\alpha^{\beta - a} f(u)\{1 - F(u)\}^{n-1}F^n(u + a)du, \end{aligned}$$

with the aid of which most of the numerical values in Table I have been computed.

5. The Limit R

The assumption involved in computing $P_2(a)$, that x_n and x_1 are independent, is equivalent to assuming that x_1 is contrasted with the greatest order statistic, y_n say, of an independently drawn sample of the same size. When a is extreme, in the sense that the probability of its being exceeded is small, it might be expected that the expressions $P_1(a)$, $P_2(a)$ will have an interesting relationship. We are led to consider the limiting value of the ratio $R(a)$ as a increases to its greatest value or indefinitely, as the case may be, that is, as $P_1(a)$ and $P_2(a)$ both converge to zero.

Consider, at first, the case where α, β are both finite. The region ω is now the triangle bounded by the lines $u = \alpha$, $v = \beta$, $v = u + a$. On applying an appropriate mean-value theorem to each of the integrals appearing in (3), we obtain the form

(6) $$R(a) = \frac{n-1}{n} \cdot \frac{\{F(v^*) - F(u^*)\}^{n-2} \int_\omega f(u)f(v)\,du\,dv}{F^{n-1}(v^{**})\{1 - F(u^{**})\}^{n-1} \int_\omega f(u)f(v)\,du\,dv},$$

where (u^*, v^*), (u^{**}, v^{**}) are points in ω. If a is allowed to approach its greatest value $\beta - \alpha$, the starred points both converge to the point (α, β), so that each of the expressions in braces in (6) converges to 1. It follows that the limit of $R(a)$ exists and is given by

(7) $$R = \frac{n-1}{n}.$$

This result, though contradicting the intuitive notion referred to above and which suggests the value 1 for R, is nevertheless distribution-free and independent of α and β. One might, therefore, be tempted to suppose that (7) would extend immediately to distributions for which one or both of α, β were not finite, on the grounds that truncation at a sufficiently remote point could have no appreciable effect on the distribution. Such a generalization, however, is false, as shown by the results in Table I.

We now sketch the outlines of a proof of the theorem that for one-tailed distributions the limit R exists and satisfies the double inequality

(8) $$\frac{n-1}{n} \leqq R \leqq 1.$$

We may, without loss of generality, take the interval of definition as $0 \leqq x < \infty$. On integrating (1) and (2) with respect to v and expanding the subsequent integrands by the binomial theorem, $R(a)$ can be expressed in the form

(9) $$R(a) = \frac{n \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^{i+1} \phi_{n-1-i,i}(a)}{n \sum_{i=1}^{n} \binom{n}{i} (-1)^{i+1} \phi_{n-1,i}(a)},$$

where

(10) $$\phi_{r,s}(a) = \int_0^\infty G^r(u) G^s(u + a) f(u)\,du$$

and

(11) $$G(u) = 1 - F(u).$$

In the successive terms of the sums in (9), with i increasing, the powers of $G(u+a)$ that occur increase while those of $G(u)$ do not. By considering the functions

(12) $$I_{r,s}{}^{(a,z)} = \frac{\int_z^\infty G^r(u)G^s(u+a)f(u)du}{\int_0^z G^r(u)G^s(u+a)f(u)du},$$

(13) $$J_{r,s}{}^{(a,z)} = \frac{\int_0^z G^r(u)G^{s+1}(u+a)f(u)du}{\int_0^z G^{r+\delta}(u)G^s(u+a)f(u)du},$$

where δ may have the value 0 or 1, it is readily shown that, uniformly in a for $0 \leqq a < \infty$,

(14) $$\lim_{z\to\infty} I_{r,s}{}^{(a,z)} = 0$$

and that

(15) $$J_{r,s}{}^{(a,z)} < G(a)/\{G(z)\}^\delta.$$

If $\delta = 1$ we may choose z such that $G(z) = \sqrt{G(a)}$, and we thus have

(16) $$\frac{\int_0^\infty G^r(u)G^{s+1}(u+a)f(u)du}{\int_0^\infty G^{r+\delta}(u)G^s(u+a)f(u)du} = J_{r,s}{}^{(a,z)} \cdot \frac{1 + I_{r,s+1}{}^{(a,z)}}{1 + I_{r+\delta,s}{}^{(a,z)}}$$

$$< \{G(a)\}^\epsilon \cdot \frac{1 + I_{r,s+1}{}^{(a,z)}}{1 + I_{r+\delta,s}{}^{(a,z)}},$$

where ϵ has the value 1 or $\frac{1}{2}$. On letting a (and z) tend to infinity, we reach the result

(17) $$R = \frac{n-1}{n} \cdot \lim_{a\to\infty} \frac{\phi_{n-2,1}{}^{(a)}}{\phi_{n-1,1}{}^{(a)}}.$$

Since $\phi_{n-1,1}^{(a)} < \phi_{n-2,1}^{(a)}$, we establish the first half of the double inequality, namely

$$R \geqq \frac{n-1}{n}. \tag{18}$$

Finally, on integrating $\phi_{n-2,1}^{(a)}$ and $\phi_{n-1,1}^{(a)}$ by parts, we obtain the relations

$$\frac{n-1}{n} \cdot \frac{\phi_{n-2,1}^{(a)}}{\phi_{n-1,1}^{(a)}} = \frac{G(a) - \int_0^\infty G^{n-1}(u) f(u+a) du}{G(a) - \int_0^\infty G^n(u) f(u+a) du} < 1, \tag{19}$$

and the second of the desired inequalities, namely,

$$R \leqq 1, \tag{20}$$

follows at once.

The result (8) is thus established.

Attempts to prove the same result for two-tailed distributions have not been successful, although the numerical results obtained suggest that it is true in general.

III. THE SUPPOSITION OF NORMALITY

6. *The Approximation $P_3(a)$*

If we suppose that x_1 and x_n are not only independent but also normally distributed, their difference $x_n - x_1$, that is, the range, is also normally distributed and, knowing the parameters of this distribution, the probability that the range will exceed any given value a can be obtained by entering the standard normal tables. This approximation to $P_1(a)$ is denoted by $P_3(a)$. The required parameters are found from the mean and variance (when they exist) of the distributions of each of the end order statistics. When the density function $f(x)$ is known, their determination is straightforward. For some of the distributions considered in Table I use has been made of the tabulated results of Hastings, Mosteller, Tukey and Winsor [5] and of Godwin [2]; for others direct calculations were made.

If μ_1, μ_n are the expected values and σ_1^2, σ_n^2 the variances of x_1, x_n, respectively, the combined assumption of independence and normality leads to the standard normal variate

$$\frac{(x_n - x_1) - (\mu_n - \mu_1)}{\sqrt{\sigma_1^2 + \sigma_n^2}} = z, \quad \text{say.} \tag{21}$$

The value of $P_3(a)$ is then given by

$$P_3(a) = \Pr\left\{z > \frac{a - (\mu_n - \mu_1)}{\sqrt{\sigma_1^2 + \sigma_n^2}}\right\}. \tag{22}$$

The results obtained by the application of this formula are set out in Table I.

7. *Application of the Approximation $RP_3(a)$*

When samples are drawn from a given population, it is necessary to have some knowledge of the percentage points of the distribution of the range in order to decide whether the range obtained in any particular case is significantly large. For this purpose we may have available data from k independent samples each of size n. If the parent population is normal and an unbiased estimate of its variance is found from these samples, the required percentage points of the distribution can be found by referring to published tables of the "Studentized" range [4]. Since it is not our purpose to try to improve on this method, any application of the approximations discussed in this paper must be to (a) cases involving non-normal distributions, or (b) problems concerned with contrasts for which no tables are available.

A survey of Table I shows that the approximation $RP_3(a)$ is likely to prove most useful for non-normal populations in the vicinity of the 5 per cent point. Restricting ourselves to those entries in the table for which $P_1(a)$ is about .05, we note first that the error made in using $RP_3(a)$ in place of $P_1(a)$ is around 16% of $P_1(a)$ in the case of normal samples (15.7% for samples of size 5, 16.5% for samples of 8 and 16.8% for samples of 10), while for the others it ranges from 3.0% (symmetrical triangular) to 65.4% (one-tailed hyperbolic). It is of interest to compare these biases with errors of estimation that may be met in practical situations.

The direct way of using the k samples of n in the non-normal case would be to compute the corresponding k ranges and determine the quantile $\omega_{.95}$, say, for this sample of k values. If this statistic is to be used for estimating the corresponding population quantile $\Omega_{.95}$, we shall need to investigate an appropriate confidence interval for $\Omega_{.95}$. The magnitude of this interval is indicated by the corresponding confidence interval for the parameter ϖ in the binomial distribution

$\binom{k}{h}\pi^h(1-\pi)^{k-h}$, where h is the number of "successes" observed in the sample of k, here the number of sample ranges greater than or equal to $\omega_{.95}$. If $k=180$ and $h=9$, the 95 per cent. confidence limits for π are .095 and .024, representing errors of 90% and 52%, respectively, if $\pi=.05$; if $k=100$ and $h=5$, the corresponding limits are .112 and .017, representing errors of 124% and 66%; if $k=60$ and $h=3$, the limits are .140 and .019, representing errors of 180% and 80%; and if $k=20$ and $h=1$, the limits are .249 and .001, representing errors of 398% and 98%. Hence the errors determined above for the approximation $RP_3(a)$ are exceeded by those obtained by using the direct method, based on the 95 per cent confidence interval, when 100 samples of 5 are available, and are extremely likely to be exceeded when even as many as 180 samples of 5 are available.

It must be recalled, however, that the errors so far associated with $RP_3(a)$ have been computed on the basis of known means and variances for the distributions of x_1 and x_n. When, as will happen in practice, the density function $f(x)$ is unknown, the relevant parameters must be estimated from the observations themselves, and the use of these estimates in place of the true values will modify the error in approximating to $P_1(a)$.

To investigate this effect, let us assume that the parent population is symmetrical and that we have available 20 independent samples of 5 observations each, so that $k=20$. $n=5$. Denote the estimates of the means and variances of x_1, x_5 by $\bar{x}_1$, $\bar{x}_5$ and s_1^2, s_5^2, respectively. No two of these estimates are strictly independent, but the dependence is probably not strong and, as we are interested here chiefly in rough comparisons, we shall assume independence. On the basis of the approximate normality of distribution of the order statistics x_1, x_5, referred to in Section 1, the standard error of $\bar{x}_5-\bar{x}_1$ is $\sigma/\sqrt{10}$ while that of both s_1^2 and s_5^2 is $\sigma^2/\sqrt{9.5}$, σ^2 being the common variance of x_1 and x_5. On replacing $\mu_5-\mu_1$ by $\bar{x}_5-\bar{x}_1$ and $\sigma_1^2+\sigma_5^2$ by $s_1^2+s_5^2$, we have, in place of (22), the approximation, $P_3^*(a)$ say, given by

$$P_3^*(a) = \Pr\left\{z > \frac{a-(\bar{x}_5-\bar{x}_1)}{\sqrt{s_1^2+s_5^2}}\right\}, \tag{23}$$

where z is $N(0, 1)$. The working approximation to $P_1(a)$ is now $RP_3^*(a)$ where, for $n=5$, $R=.8$ for all distributions defined over finite intervals and may presumably be taken as .8 for all two-tailed distributions that do not depart widely from normality.

In the special case of a normally distributed parent population we have, quoting from [2],

$$\mu_5 = -\mu_1 = 1.16296, \qquad \sigma_1^2 = \sigma_r^2 = .447535,$$

giving, for $a=3.86$, $P_1(a)=.0498$ and $P_3(a)=.0525$. If $\bar{x}_5-\bar{x}_1$ overestimates $\mu_5-\mu_1$ by one standard error, say, and simultaneously σ_1^2, σ_5^2 are each overestimated by one standard error, we find $P_3^*(a)=.1122$ and $.8P_3^*(a)=.0898$. We have taken one standard error in each case rather than the customary two, which correspond roughly to the 95 per cent confidence limits, but we have supposed that the errors accumulate whereas they will often tend to cancel. Even though the result .0898 overestimates $P_1(a)$ by about 80%, it compares quite favourably with the upper 95 per cent confidence limit obtained by the direct method. A similar procedure on the side of underestimation gives $.8P_3^*(a)=.0099$, as compared with .001, the lower 95 per cent confidence limit obtained by the direct method.

The above conjectures have been tested by means of random drawings of samples of 5 from (*i*) the standard normal distribution, using the tables of random deviates given in [7]; (*ii*) the rectangular distribution $f(x)=1$ for $0\leq x\leq 1$, using the five-figure tables of random numbers given in [1] and [9]; and (*iii*) the beta distribution for the values 2, 2 of the parameters, namely $f(x)=6x(1-x)$ for $0\leq x\leq 1$, using five-figure tables of random numbers in conjunction with the t-tables for 4 degrees of freedom given in [4] and subsequently transforming the variate values found. The results are collected in Table II, corresponding to the values 20, 60, 100 and 180 of k and to the value .8 of R. (In each case, the 180 samples are the compound of all the preceding ones.) The values of the relevant parameters are also included, corresponding to the value ∞ of k.

It is observed that for the normal and beta distributions the approximation $RP_3^*(a)$ is extremely good for all values of k. For the rectangular distribution it is only slightly less satisfactory. In only three situations is the difference between $\mu_5-\mu_1$ and $\bar{x}_5-\bar{x}_1$ of the same sign as that between $\sigma_1^2+\sigma_5^2$ and $s_1^2+s_5^2$. The greatest percentage error in $P_3^*(a)$ relative to $P_3(a)$ on the side of overestimation is 46.7 (rectangular for $k=20$) and on the side of underestimation 53.8 (beta for $k=20$). In no case have errors as large as those conjectured above for the normal case (114% on the side of overestimation, 77% on that of underestimation) been approached.

We conclude that the use of the approximation $[(n-1)/n]P_3^*(a)$ is, for samples of small size, preferable to the direct method when the number of samples available lies between 20 and 180.

TABLE II

RESULTS OF RANDOM DRAWINGS—CONTRAST x_5-x_1

Distribution	k	$\bar{x}_1$	$\bar{x}_5$	s_1^2	s_5^2	$.8P_1^*(a)$	Standardised overestimation of $\mu_5-\mu_1$	Standardised overestimation of $\sigma_1^2+\sigma_5^2$	Percentage error in $P_1^*(a)$ relative to $P_1(a)$
Rectangular	20	.1138	.8416	.01861	.01555	.1193	− .88	− .86	+46.7
$a=.92$	60	.1420	.8354	.01384	.01710	.0692	+1.04	−2.38	−14.9
$P_1(a)=.0544$	100	.1516	.8407	.01773	.02060	.0953	+1.12	− .48	+17.1
$.8P_1(a)=.0814$	180	.1442	.8390	.01648	.01869	.0919	+1.89	−2.14	+13.0
	∞	.1667	.8333	.01984	.01984				
Beta	20	.2272	.7686	.00543	.01150	.0266	+ .51	−2.83	−53.8
$a=.78$	60	.2235	.7698	.01104	.01603	.0622	+1.09	−1.47	+ 8.1
$P_1(a)=.0497$	100	.2417	.7434	.02077	.01422	.0547	−1.11	+1.66	− 5.0
$.8P_1(a)=.0576$	180	.2340	.7550	.01578	.01454	.0547	− .03	− .59	− 5.0
	∞	.2393	.7607	.01565	.01565				
Normal	20	−1.177	1.108	.4228	.7307	.0570	− .20	+1.85	+35.8
$a=3.86$	60	−1.169	1.203	.5054	.3849	.0459	+ .38	− .06	+ 9.3
$P_1(a)=.0498$	100	−1.023	1.115	.5046	.3822	.0270	−1.98	− .13	−35.8
$.8P_1(a)=.0420$	180	−1.089	1.144	.4959	.4176	.0354	−1.31	+ .41	−15.6
	∞	−1.163	1.163	.4475	.4475				

IV. EXTENSIONS TO OTHER CONTRASTS

8. *The Contrast* $x_n-\frac{1}{2}(x_1+x_2)$

The general procedure outlined above can be applied to contrasts other than the range. As a simple extension we shall consider the contrast $x_n-\frac{1}{2}(x_1+x_2)$. The expression for $P_1(a)$ is obtained from the joint probability density function of the three order statistics written, upon integration over the region for which the inequalities $x_2>x_1$, $x_n>\frac{1}{2}(x_1+x_2)$, $x_n>x_2$ are simultaneously satisfied. The integrand is here

$$n(n-1)(n-2)f(x_1)f(x_2)\left\{F(x_n)-F(x_2)\right\}^{n-3}f(x_n). \tag{24}$$

Assuming that $f(x)$ is defined over the interval $(-\infty, \infty)$ and that the order of integration is x_n, x_2, x_1 in turn, the integral is most easily evaluated as the sum of two triple integrals for which the terminals are, respectively in order,

$$\left(\frac{x_1+x_2}{2}+a,\ \infty\right),\quad (x_1,\ x_1+2a),\quad (-\infty,\ \infty)$$

and (x_2, ∞), (x_1+2a, ∞), $(-\infty, \infty)$. When the density function is defined over a finite interval (α, β), the integral is evaluated as a triple integral for which the terminals are, in order,

$$\left(\frac{x_1 + x_2}{2} + a, \beta\right), (x_1, 2\beta - 2a - x_1), \quad (\alpha, \beta - a).$$

In forming the approximation $P_2(a)$ we assume that x_1 and x_2 are correlated but are both independent of x_n, leading to the integrand

$$n^2(n-1)f(x_1)f(x_2)\{1 - F(x_2)\}^{n-2}F^{n-1}(x_n)f(x_n). \tag{25}$$

When $f(x)$ is defined over the finite interval (α, β) it is readily shown that the ratio $R(a) = P_1(a)/P_2(a)$ converges to the limit

$$R = (n-2)/n, \tag{26}$$

as a converges to $\beta - \alpha$. No attempt has been made to investigate bounds for R when $f(x)$ is defined under more general conditions, although it is conjectured that the above limit is applicable to the normal case and that, in general, the upper bound is 1.

In investigating the approximation $P_3(a)$, which is based on the assumption that the variates $\frac{1}{2}(x_1+x_2)$ and x_n are normally and independently distributed, we require the means and variances of x_1, x_2, x_n as well as the covariance between x_1 and x_2. If these parameters are, in turn, μ_1, μ_2, μ_n and σ_1^2, σ_2^2, σ_n^2, σ_{12}, the present assumption leads to the standard normal variate

$$z = \frac{[x_n - \frac{1}{2}(x_1 + x_2)] - [\mu_n - \frac{1}{2}(\mu_1 + \mu_2)]}{\sqrt{\frac{1}{4}(\sigma_1^2 + \sigma_2^2 + 2\sigma_{12}) + \sigma_n^2}}, \tag{27}$$

so that, for given $a > 0$,

$$P_3(a) = \Pr\left[z > \frac{a - [\mu_n - \frac{1}{2}(\mu_1 + \mu_2)]}{\sqrt{\frac{1}{4}(\sigma_1^2 + \sigma_2^2 + 2\sigma_{12}) + \sigma_n^2}}\right]. \tag{28}$$

The indicated approximations to $P_1(a)$ are thus $[(n-2)/n]P_2(a)$ and $[(n-2)/n]P_3(a)$.

If the density function $f(x)$ is unknown, use may be made of the last-written formula in approximating to $P_1(a)$, where the parameters appearing on the right-hand side of the expression (28) for $P_3(a)$ are to be estimated. As before, we may have available k samples of n for this purpose. When the appropriate estimates are inserted, the corresponding probability may be denoted by $P_3^*(a)$, and for distributions that are defined over a finite interval the working approximation to $P_1(a)$ then becomes $[(n-2)/n]P_3^*(a)$. The same rule may be presumed for distributions that are normal or nearly normal.

Some of the above considerations have been applied to samples of 5 from populations previously considered, namely the rectangular dis-

tribution, the symmetrical triangular distribution, the beta distribution with parameters 2, 2, and the normal distribution.

For the first of these we have $f(x)=1$ for $0 \leqq x \leqq 1$ and the integrals defining $P_1(a)$, $P_2(a)$ are readily resolved. To find $P_3(a)$ we used low moments given in [5].

For the triangular distribution $f(x)=x$ for $0 \leqq x \leqq 1$ and $f(x)=2-x$ for $1 \leqq x \leqq 2$, the evaluation of the integrals tends to become tedious and, accordingly, $P_2(a)$ was not computed. For the beta distribution with parameters 2, 2, $f(x)=6x\ (1-x)$ for $0 \leqq x \leqq 1$, the evaluation of $P_1(a)$, though exceedingly laborious, was carried through for three values of a.

For the standard normal distribution, $P_1(a)$ was evaluated by the method of double quadrature, using Simpson's rule. Here $P_1(a)$ was first expressed in the form

$$P_1(a) = 1 - 20 \int_{-\infty}^{\infty} \int_{-a}^{a} f(x) f(a + x + u) \{F(\tfrac{3}{2}a + x + \tfrac{1}{2}u) - F(a + x + u)\}^3 du\, dx \tag{29}$$

and intervals of 0.2 were chosen for both u and x.

The results are collected in Table III. It will be observed that when $P_1(a)$ lies between .01 and .05, $RP_3(a)$ affords an extremely close approximation to $P_1(a)$ for the triangular, beta and normal distributions, while for the rectangular distribution the approximation is excellent

TABLE III

TAIL PROBABILITIES FOR THE DISTRIBUTION OF $x_3 - \frac{1}{2}(x_1 + x_2)$

Distribution	a	$P_1(a)$	$P_2(a)$	$P_3(a)$	$R(a)$	R	$RP_3(a)$
Rectangular $f(x)=1$ for $0 \leqq x \leqq 1$	.848	.0507	.0704	.0947	.720	.6	.0568
	.900	.0162	.0239	.0582	.678	.6	.0349
	.915	.0103	.0154	.0500	.667	.6	.0300
Symmetrical triangular $f(x)=\begin{cases} x \text{ for } 0 \leqq x \leqq 1 \\ 2-x \text{ for } 1 \leqq x \leqq 2 \end{cases}$	1.303	.0503		.0720		.6	.0432
	1.480	.0105		.0228		.6	.0137
Beta, parameters 2, 2 $f(x)=6x\ (1-x)$ for $0 \leqq x \leqq 1$	.68	.0621		.0907		.6	.0544
	.70	.0452		.0731		.6	.0438
	.75	.0184		.0404		.6	.0242
Standard normal $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ for $-\infty < x < \infty$						(?)*	(?)*
	3.0	.1004		.1221		.6	.0733
	3.2	.0676		.0815		.6	.0489
	3.4	.0464		.0519		.6	.0312
	4.0	.0094		.0102		.6	.0061

* Question marks here have the same significance as in Table I.

when $P_1(a)$ is of the order of .05 and only slightly astray for smaller values of $P_1(a)$.

The approximations $P_3(a)$ are based on known values of the parameters in the distributions of the indicated order statistics. As mentioned above, these parameters will, in practice, be estimated from the data provided by groups of samples. To investigate the effect of this procedure on the working approximation to $P_1(a)$, samples of 5 were drawn from the beta and normal distributions, in manners previously described, and the values of x_1, x_2, x_5 were listed. For the beta distribution, three sets of 20 samples were drawn randomly from an original lot of 100 randomly selected samples; for the normal distribution one set of 20 samples and four independent sets of 100 samples were drawn.

The results of these drawings are collected in Table IV for the values of k mentioned, together with the values of $P_1(a)$ and the working approximation $.6P_3{}^*(a)$ for selected values of a. For convenience in making comparisons, the table includes the values of the parameters themselves and of the approximation $.6P_3(a)$, corresponding to the value ∞ of k.

TABLE IV

RESULTS OF RANDOM DRAWINGS—CONTRAST $x_5-\frac{1}{2}(x_1+x_2)$

Distribution	k	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_5$	s_1^2	s_2^2	s_{12}	s_5^2	$.6P_3^*(a)$		
Beta									$a=.68$	$a=.70$	$a=.75$
$P_1(.68)=.0621$	20	.2812	.4157	.7060	.03180	.02402	.02293	.01105	.0274	.0219	.0120
$P_1(.70)=.0452$	20	.2712	.4791	.7646	.01632	.01347	.00826	.01185	.0173	.0127	.0056
$P_1(.75)=.0184$	20	.2464	.4305	.7847	.01572	.01933	.01309	.01303	.0494	.0395	.0213
	100	.2417	.4097	.7434	.02077	.02367	.01524	.01422	.0455	.0360	.0201
	∞	.2393	.3786	.7607	.01565	.01760	.01040	.01565	.0544	.0438	.0242
Normal									$a=3$	$a=3.4$	$a=4$
$P_1(3)=.1004$	20	−1.049	−.391	1.280	.4396	.2269	.1365	.4229	.0653	.0253	.0041
$P_1(3.4)=.0464$	100	−1.289	−.559	1.136	.4852	.3305	.2528	.4754	.0884	.0406	.0092
$P_1(4)=.0094$	100	−1.234	−.499	1.243	.5314	.3394	.2428	.5578	.1042	.0519	.0138
	100	−1.228	−.478	1.190	.4750	.3085	.2396	.4831	.0853	.0387	.0086
	100	−1.228	−.483	1.135	.3951	.3246	.1843	.3963	.0651	.0254	.0042
	∞	−1.163	−.495	1.163	.4475	.3115	.2243	.4475	.0733	.0312	.0061

In the first set of 20 drawings from the beta distribution the error in $P_3{}^*(a)$ relative to $P_3(a)$ is, for each value of a, of the order of 50%; for the second set of 20 it is of the order of 73%; for the third set it is of the order of 10%; while the average of the three sets is of the order

of 44%. For the original set of 100 drawings the error is of the order of 17%.

For the normal samples, the set of 20 drawings leads to errors of 11% when $a=3$, 19% when $a=3.4$ and 33% when $a=4$. The means of the errors for the four sets of 100 drawings are 23% when $a=3$, 35% when $a=3.4$ and 62% when $a=4$.

One set of 20 from the beta distribution yields low percentage errors while the other two have high values. The errors exhibited by the single set of 20 samples from the normal distribution appear to be exceptionally low. The first three sets of 100 drawn each display higher errors, the second being, perhaps, exceptionally high. The fourth set has errors almost identical with those of the independent set of 20.

From these considerations it appears that about 100 samples of 5 would seem to be required in order that the approximation $.6P_3^*(a)$ should provide a reliable test criterion.

As in Section 6, the direct method of procedure would lead to errors of 124% and 66%, corresponding to the 95 per cent confidence limits for π when $\pi=.05$, $k=100$, $h=5$. Upon examination of Table IV we find that for the beta distribution, with $k=100$ and $a=.7$, the percentage error in $.6P_3^*(a)$ relative to $P_1(a)$ is 20% on the side of underestimation. For the normal distribution, with $k=100$ and $a=3.4$, the errors vary from 12% on the side of overestimation to 45% on the side of underestimation.

We conclude that the use of the approximation $[(n-2)/n]P_3^*(a)$ is preferable to the direct method in the circumstances described.

9. *Concluding Remarks*

For any contrast it is possible to write down immediately the expressions for $P_1(a)$ and $P_2(a)$, the latter being based on convenient assumptions of independence between the groups of order statistics that enter into the contrast. When, as in all practical situations, $f(x)$ is defined over a finite interval (α, β), the calculation of R is straightforward.

Thus, considering the contrast $\frac{1}{2}(x_{n-1}+x_n)-\frac{1}{2}(x_1+x_2)$ and assuming that the end pairs of order statistics are independent of each other in forming the approximation $P_2(a)$, we find, for any distribution defined over a finite interval,

$$R = \frac{(n-2)(n-3)}{n(n-1)},$$

while, for the contrast $x_n-\frac{1}{3}(x_1+x_2+x_3)$ we find, under like conditions,

$$R = \frac{n-3}{n}.$$

With the aid of such coefficients working approximations to the true probability $P_1(a)$, involving the principle of normality, can be developed generally by direct extensions of the methods described above.

10. *Acknowledgments*

Our thanks are due to Professor J. W. Tukey, who drew our attention to the original problem and who offered critical and constructive advice during the progress of the study. We are indebted to Mr. H. J. Godwin for some suggestions regarding the evaluation of $P_2(a)$ for the range of the standard normal distribution, and to Professor E. R. Love for a proof of the limit $(n-1)/n$ for R for the general chi-square distribution when m increases indefinitely. For assistance with the vast amount of computing that lies concealed behind the tabular entries throughout the paper our special thanks are due to Miss Alison Doig.

REFERENCES

[1] Fisher, R. A., and Yates, F., *Statistical Tables etc.*, Edinburgh, Oliver and Boyd, 1942.

[2] Godwin, H. J., "Some Low Moments of Order Statistics," *Annals of Mathematical Statistics*, 20 (1949), 279–85.

[3] Hald, A., *Statistical Theory with Engineering Applications*, New York, John Wiley and Sons, 1952, p. 320.

[4] Hartley, H. O., and Pearson, E. S., "Tables of the Probability Integral of the *t*-Distribution," *Biometrika*, 37 (1950), 168.

[5] Hastings, C., Jr., Mosteller, F., Tukey, J. W., and Winsor, C. P., "Low Moments for Small Samples: A Comparative Study of Order Statistics," *Annals of Mathematical Statistics*, 18 (1947), 413–26.

[6] Kendall, M. G., *The Advanced Theory of Statistics*, v. 1, London, Charles Griffin and Co., (1943), p. 221.

[7] Mahalanobis, P. C., ed., "Tables of Random Samples from a Normal Population, *Sankhyā*, 1 (1934), 289–328.

[8] Pearson, E. S., and Hartley, H. O., "Tables of the Probability Integral of the 'Studentized' Range," *Biometrika*, 33 (1943) 89.

[9] "Random Digits" *Journal of the American Statistical Association*, 47 (1952), 710.

OPTIMUM GROUPING IN ONE-CRITERION VARIANCE COMPONENTS ANALYSIS

E. P. KING*
National Bureau of Standards

TESTS of significance associated with the single criterion analysis of variance usually assume that a sample of n observations is drawn from each of m normal populations with common variance σ^2. In the "components of variance" model, the m population means are themselves considered a sample of m observations on a superpopulation, also normal, with variance $\theta^2\sigma^2$. The null hypothesis $\theta=0$ is tested against the alternative $\theta>0$ by means of the F ratio. Detailed descriptions of this model are given by Eisenhart ([3] and [4]); Ferris, Grubbs, and Weaver [5]; and Crump [2].

When the number of populations is indefinite and the total number of observations $N=mn$ is limited, it is possible to determine which (m, n) combination gives the most powerful F-test for a given N. This problem was considered in [1], [4], and [5]. For all cases examined, each (m, n) combination in turn was found to provide the most powerful test for some interval of θ.

An extensive list of these optimum groupings would seem, however, of limited practical interest; for the applied statistician is seldom in a position to specify in advance the magnitude of θ which he desires to detect. To remove the need of *a priori* information, this note proposes a simple rule for selecting m and n which gives nearly maximum power for all $\theta>0$.[1] The rule is to select m and n as nearly equal as possible. The operating characteristics obtained from this selection are shown in Figure 1 for $m=n=8$, 10, 12, and 16 when the test is conducted at the 5 per cent level of significance. The values used to plot the curves were obtained for the most part by the well-known method outlined in [5] using the percentiles of the F-distribution as given in [6]. Table 8.4 in [4] also served as a valuable check on the results. The dashed curves enclose the operating characteristics for all the (m, n) combinations that can be formed from the given amount of data. These curves approximate the upper and lower envelopes for the family; no single (m, n) choice will yield either curve as its operating characteristic.

Groupings in which m is much less than n are more powerful than

* Now with Eli Lilly and Company.

[1] It should be emphasized that this paper deals wholly with significance tests rather than estimation. If one is interested in estimating θ or $\theta^2\sigma^2$, the suggested procedure may not be at all optimum.

those for which m is approximately equal to n for very small values of θ, but in this range *no* grouping gives a test of sufficient power for practical use. Choosing m much greater than n gives a slightly more powerful test for very large values of θ, but in this range the choice $m=n$ gives power very close to unity.

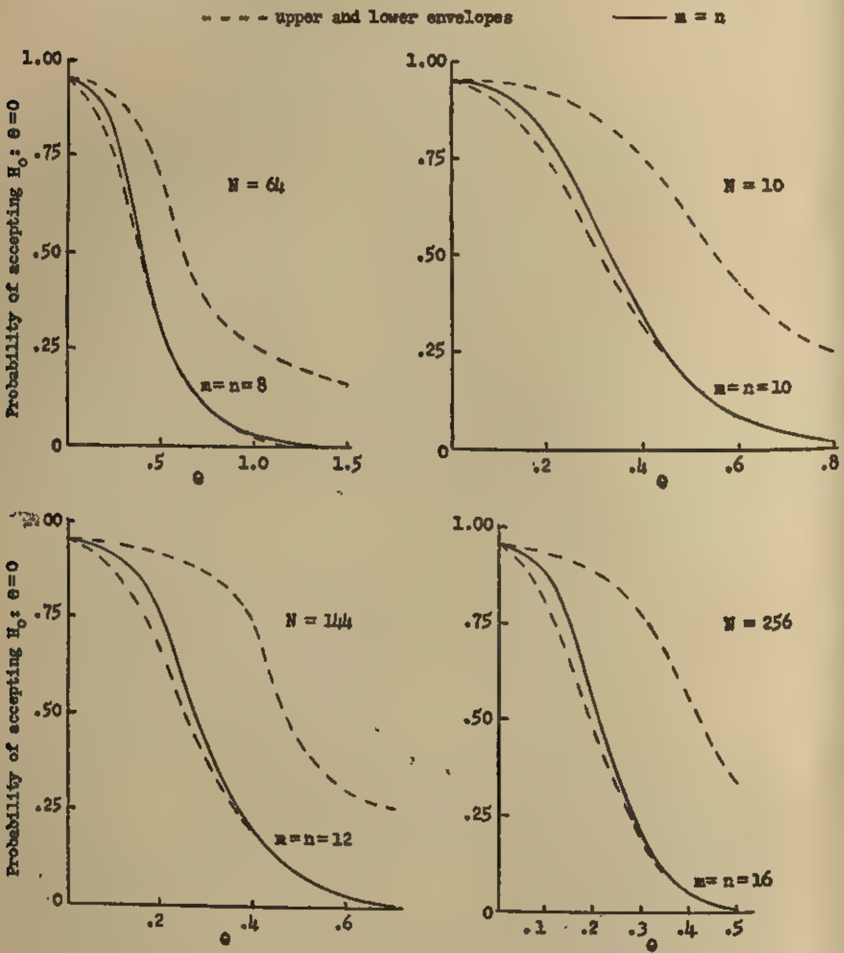


FIG. 1. Operating characteristics of the F-test for testing $\theta=0$ against $\theta>0$ at the 5% level of significance.

REFERENCES

[1] Baines, A. H. J., "On economical design of statistical experiments," (British) Ministry of Supply, Advisory Service on Statistical Methods and Quality Control, *Technical Report, Series R, No. Q.C./R/15*, July 15, 1944.

[2] Crump, S. Lee, "The present status of variance component analysis," *Biometrics* (1951), 1–16.

[3] Eisenhart, Churchill, "The assumptions underlying the analysis of variance," *Biometrics*, 3 (1947) 1–21.

[4] Eisenhart, Churchill, "Planning and interpreting experiments for comparing two standard deviations," Chapter 8 in Statistical Research Group, Columbia University, *Techniques of Statistical Analysis*, edited by Churchill Eisenhart, Millard W. Hastay, and W. Allen Wallis, New York, McGraw-Hill Book Co., 1947, 267–314.

[5] Ferris, C. D., Grubbs, F. E., and Weaver, C. L., "Operating characteristics for common statistical tests of significance, "*Annals of Mathematical Statistics*, 17 (1946) 178–97.

[6] Hald, A., "Fractiles of the v^2-Distribution," Table VII in *Statistical Tables and Formulas*, New York, John Wiley and Sons, 1952, 47–59.

STATISTICAL ABSTRACTS

All communications concerning this section should be addressed to the Abstracts Editor, Professor George E. Nicholson, Jr., Chairman of the Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Anderson, T. W., "On estimation of parameters in latent structure analysis," *Psychometrika*, 19 (1954), 1–10.

A new computational procedure for estimating the parameters in latent structure analysis is developed. The procedure has the advantage of avoiding the use of implicitly defined and unobservable quantities as well as being relatively simple computationally. On the other hand the proposed procedure has the disadvantage of using only part of the available information and of using that part asymmetrically. A numerical example is worked out in detail.

After reviewing the basic model of latent class structure, the author develops the rationale for the estimation method suggested as the basis for computation. The method deduces the proportion of the population in each latent class, and the probabilities of positive responses for each individual in the latent classes, from knowledge of probabilities of positive responses from individuals in the population as a whole. There are several possible choices of initial data for estimating the same parameters—one may obtain differing estimates for the same parameters depending upon the particular initial choice made. If the latent classes are well defined, however, the range of the differences between equivalent estimates will be relatively small. B. J. Winer, *University of North Carolina.*

Bartlett, M. S., and Rajalakshman, D. V., "Goodness of fit tests for simultaneous autoregressive series," *Journal of the Royal Statistical Society*, (B) 15 (1953), 107–24.

"The simplified method of derivation of Quenouilles (1947) goodness of fit test for autoregressive series with discrete time, . . . , is extended to the case of simultaneous autoregressive series . . . . The general solution is applied to a particular first-order process in two variables." G. L. Edgett, *Virginia Polytechnic Institute.*

Beall, Geoffrey, and Rescia, Richard R., "A generalization of Neyman's contagious distributions," *Biometrics*, 9 (1953), 354–86.

A class of discrete distributions depending on three parameters, one of which is called n, is constructed. It is shown that Neyman's contagious distributions of types A, B, and C are members of this class for $n=0, 1, 2$ respectively. The calculation of the probabilities in these distributions require the use of recursive relations which are presented here. The estimation of the parameter n is a problem which the authors treat by fitting the frequency of cases with zero occurrences; the remaining two parameters are estimated by the method of moments. Several examples are discussed where it is shown that values of n other than 0, 1 or 2 give better fit than is obtained with the distributions previously employed. Lincoln Moses, *Stanford University.*

Bechhofer, Robert E., "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Annals of Mathematical Statistics*, 25 (1954), 16–39.

The problem of classifying normal populations in two or more groups with respect to the ranking of their true means has been considered. The lower bounds for the probability of correct grouping is obtained by considering the least favourable configuration of the means consistent with the given rankings. In the case of two groups, the least favourable configuration of the means is $\mu_{[1]}=\cdots=\mu_{[k-t]}$ for the lower group and $\mu_{[k-t+1]}=\cdots=\mu_{[k]}$ for the upper group and the tables of probabilities of correct grouping have been constructed for values of k and t and

$$\lambda=\frac{\mu_{[k-t+1]}-\mu_{[k-t]}}{\sigma}$$

where σ is the common standard deviation of the populations. These tables could conversely be used to determine the sample sizes when one wants to ensure a certain lower bound of the probability of correct grouping for $\lambda>\lambda_0$. The method has been generalised to two-way classifications when the mean of the variable X_{ij} is given by $\mu+\alpha_i+\beta_j$, and the experimenter wants to pick up the population with highest α_i, and highest β_j. Illustrations of the use of the tables have been given. M. N. Ghosh, *University of North Carolina.*

Birnbaum, Allan, "Admissible tests for the mean of a rectangular distribution," *Annals of Mathematical Statistics*, 25 (1954), 157–61.

The problem of testing for the mean θ of a rectangular distribution has been considered when the range is known and when the loss function is simple, i.e. assumes the values zero and one for correct and incorrect decisions respectively. Using the fact that the minimum observation u, and the maximum observation v are joint sufficient statistics, the Bayes solutions are obtained in terms of functions of u and v. Using the Neyman-Pearson Lemma, the most powerful one-sided and two-sided tests are obtained from the class of Bayes solutions. M. N. Ghosh, *University of North Carolina.*

Box, G. E. P., and May, W. A., "A statistical design for efficient removal of trends occurring in a comparative experiment with an application in biological assay," *Biometrics*, 9 (1953), 304–19.

It is desired to compare the dose response curves of two biological preparations. A trend in time is expected to exist, which would ordinarily be confounded with the dosage response relations. An example involving four dose levels, each repeated for both drugs is given in full detail. The dose levels to be given on the eight occasions (equally spaced in time) are so chosen that linear and quadratic dose effects, linear, quadratic and cubic time trend effects, and interactions between drugs and each of the five named effects can all be estimated and all estimates be mutually orthogonal. Since the dose response curves turn out as parallel straight lines in the example, and the dose metameter is log dose, relative potency is estimated and a confidence interval is given. The last section of the paper deals with the method used to choose dose levels yielding the desired properties of the design. Lincoln Moses, *Stanford University.*

Chakravarti, N., and Bandyopadhyay, K. S., "A note on the consumption of cereal per adult unit in Calcutta," *Sankhya*, 13 (1953), 215–18.

Survey information from a family budget inquiry conducted during 1950–51 by the State Statistical Bureau, Government of West Bengal, utilizing only the data for Calcutta, forms the basis of this note. The method of least squares is used to estimate the consumption of cereal in a given age or sex group of a random sample for the town of Calcutta, India. T. S. Russell, *Virginia Polytechnic Institute.*

Chapman, Douglas G., "The estimation of biological populations," *Annals of Mathematical Statistics*, 25 (1954), 1–15.

This paper gives a systematic review of the various methods of sampling used for estimation of wild population. The mathematical models and assumptions are explicitly stated, which serves a very useful purpose of forewarning the consumers of these statistical methods about the pitfalls in this area. Most of these sampling methods depend on tag-recapture technique, which could possibly be useful for estimating fish population. Other interesting methods depending on size of successive samples developed by DeLury and others are also discussed. M. N. Ghosh, *University of North Carolina.*

Dwyer, P. S., "Solution of the personnel classification problem with the method of optimal regions," *Psychometrika*, 19 (1954), 11–26.

The mathematical problem involved in personnel classification is shown to be a special case of the general mathematical problem of linear programming. After presenting numerical examples of variations in the general problem that arise in the area of personnel classification, the author points out that equivalent problems are encountered in the Hitchcock transportation problem, problems that arise in biometric classification, and problems encountered in a zero-sum two-person game. Essentially the classification problem is that of finding coefficients that maximize a linear form, subject to a set of linear constraints.

For most practical purposes special computational procedures rather than the more general mathematical solution seem most feasible. The conditions underlying the method of optimal regions as developed by the author are generalizations of those given earlier by Brogden. Computationally the method is an iterative one which basically moves hyperplanes parallel to successive positions in such a way that the optimal solutions, involving the number of points within the resulting regions, eventually satisfy the desired quotas. In most practical problems encountered so far, the method leads to a solution in relatively few iterations. It is recommended for use for problems in which there are preassigned quotas and only a small number of categories. In application to a problem in which there were 1152 men and 7 job categories, a solution was attained after eight iterations.

This article provides an excellent and readable summary of work that has been

done in this area. B. J. WINER, *University of North Carolina.*

Grubbs, Frank E., and Coon, Helen J., "On setting test limits relative to specification limits," *Industrial Quality Control*, 10 (1954), 15.

Specification limits for a product should not be used as limits for determining the acceptability of the product if errors of measurement will be made in the testing. The paper shows how to determine test limits which satisfy several different criteria. If we let A be the chance of and C_A the cost of accepting a non conforming piece and B be the chance and C_B the cost of rejecting a conforming piece, then the total expected cost of making wrong decisions is C_AA+C_BB.

A general expression for the test limits which minimize this expected cost is presented and tables of factors which determine the limits are given for $C_A=C_B$, $C_A=2C_B$, and for $A=B$.

A major conclusion of the paper is that in most cases the test limits should be outside the specification limits unless $C_A\geqq 6C_B$. In the analysis it is assumed that the product quality and the measurement errors are normally distributed with known variances. ALBERT H. BOWKER, *Stanford University.*

Gumbel, E. J., "The maxima of the mean largest value and of the range," *Annals of Mathematical Statistics*, 25 (1954), 76–84.

In this article the author generalizes some earlier results obtained by both Plackett and Moriguti. He shows that the maximum of the mean range calculated by Plackett holds for any continuous variate possessing the first two moments.

The mean and the standard deviation of the largest value and the mean range are given for a distribution where the mean largest value is a maximum, and for another distribution where the mean range is a maximum.

The asymptotic properties of the reduced values for the two distributions are compared.

Graphs for probability and density functions which maximize the mean largest value ($n=2,3,4,5$) and for the mean largest reduced values and mean ranges as functions of n are drawn. A. E. SARHAN, *University of North Carolina.*

Guttman, L., "Image theory for the structure of quantitative variates," *Psychometrika*, 18 (1953?, 277–96.

A new structural model which purports to avoid problems of indeterminancy inherent in the factor analysis model is presented. Within a universe of variates, each variate is partitioned into a part that is predictable by linear multiple regression from the other variates (a common part) and a part that is not predictable (an alien part). Sampling of variates within this universe defines partial images for each variate (i.e., predicted values from $n-1$ remaining variates); the non-predictable parts are called partial anti-images. The author contends that this partition is unique, whereas the factor analysis model provides no unique definition for the partitioning of the variates. The factor analysis model is said to be determinate only if it reduces to the image analysis model as the size of the sampling from the universe increases.

Under the partition made by image analysis, the correlation between two variates can be expressed as the difference between the covariance of the partial images and the covariance of the partial anti-images. Using this identity computational procedures for structural analysis of interrelationships are developed. It is pointed out that in the special case of determinate common factors, non-diagonal covariances of partial anti-images must approach zero. B. J. WINER, *University of North Carolina.*

Harman, H. H., "The square root method and multiple group methods of factor analysis," *Psychometrika*, 19 (1954), 39–55.

Multiple group methods of factoring correlation matrices have been used for somewhat varying purposes by various authors. By some it is considered as a convenient method for extracting factors, by others it is considered as a means for locating reference vectors in accordance with a predetermined hypothesis. The author attempts to integrate the various approaches to multiple group factoring methods and develops a systematic notation for these methods.

Phases of the computational procedures are shown to be simplified by application of the square root method for solving sets of simultaneous linear equations. Basically the square root method is a computational technique for factoring a matrix into two triangular matrices. Detailed steps in the numerical application of the square root method are given. Its application to multiple group factor analysis, computation of inverses, and regression analysis in general are clearly and compactly presented. For many purposes the square root algorithm is shown to be superior to the Doolittle al-

gorithm. B. J. WINER, *University of North Carolina.*

Hartley, H. O., and David, H. A., "Universal Bounds for Mean Range and Extreme Observation," *Annals of Mathematical Statistics*, 25 (1954), 85–99.

The range, mean range, and other recent techniques of short-cut analysis of variance are used in industrial quality control under the assumption that the basic distribution is normal. For example, an observed range may be converted to an unbiased estimate of the standard deviation by multiplication with a certain constant. The authors consider to what extent this estimate is biased when the basic distribution is not normal. The problem of establishing upper bounds for E (range)/σ has been considered by both Plackett and Moriguti. They tabulated upper bounds but little was done with the lower bound. The authors show that Moriguti's solution, which is confined to finding an upper bound for symmetrical distributions, applies in general. Also, the authors derive universal upper and lower bounds for the ratio E (range)/σ for any $f(x)$ in which $a \leq x \leq b$, (a and b are constants). Universal upper bounds are given for $E(X_n)/\sigma$ for the case where X is finite and X_n is the largest sample element. A. E. SARHAN, *University of North Carolina.*

Jowett, G. H., and Scott, J. F., "Simple graphical techniques for calculating serial and spatial correlations and mean semi-squared differences," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 81–86.

Two techniques for getting the approximate value of the serial correlation are described. The value of the Mean Semi-squared Difference (MSSD) $= \frac{1}{2}$ Average $(x_i - x_{i+s})^2$ is determined by a "tracing" method and a "transparent scale" method. The approximate value of the serial correlation is then [Average $(x_i - \bar{x})^2 -$ MSSD]/Average $(x_i - \bar{x})^2$. The graphical methods compare favorably with direct use of a desk calculator. While they are less accurate, they are less tedious, can be carried out with simple equipment, and the tracing method is actually much quicker when statistics are required for a large number of logs. PAUL N. SOMERVILLE, *Virginia Polytechnic Institute.*

Kamat, A. R., "Some properties of estimates for the standard deviation based on deviations from the mean and variate differences," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 233–40.

Estimates of the relative variances of Σ_{p0} and Σ_{pr} are given, where Σ_{p0} and Σ_{pr} are defined as the unbiased estimates of σ^p based respectively on the pth power of the deviation from the mean, and the pth power of the absolute values of $\Delta^r x_i$, the rth order variate differences. The relative variance is defined as the square of the coefficient of variation (Fisher). Putting $S_{pr} = (\Sigma_{pr})^{1/p}$, it is shown that for large samples the relative variance of S_{2r} is smaller than that of S_{pr} if $p \neq 2$ and that the relative variance of S_{3r} is less than that of S_{1r}. Some results are given for small samples. If the values in the sample have a slight linear trend in their means, it is shown that Σ_{11} and Σ_{21} have a much smaller bias than Σ_{10} or Σ_{20}. Further, the increase in their variance is much smaller. If $\omega_{pr} = \Sigma_{pr}/s^p$ where $s^2 = \Sigma_{20}$, then it is proved that the kth moment about zero of ω_{pr} is equal to the ratio of the kth moments about zero of the numerator and denominator of ω_{pr}. PAUL N. SOMERVILLE, *Virginia Polytechnic Institute.*

Kempthorne, O., and Tischer, R. G., "An example of the use of fractional replication," *Biometrics*, 9 (1953), 295–303.

It was desired to study the effect of various factors on the acceptability of dehydrated corn. The relevant factors selected for study in an exploratory experiment were: varieties (8), date of harvesting (4), blanching condition (2), temperature of dehydration (2), temperature of storage (2), length of storage (2). In addition it was considered desirable to use 4 blocks rather than a completely randomized design. A complete experiment would thus require $2048 = 2^{11}$ plots. Both blocks and varieties could be represented as pseudo-factors in a 2^k factorial design, so a $\frac{1}{4}$ replicate was chosen which resulted in no 2-factor interactions being mutually confounded. In addition the design had some features of a split-split-block design.

The rationale for selection of the design is fully presented and the analysis and results are sketched. LINCOLN MOSES, *Stanford University.*

Lancaster, H. O., "A reconciliation of χ^2, considered from metrical and enumerative aspects," *Sankhya*, 13 (1953), 1–10.

The relationship between the χ^2 of the goodness of fit test and the deviations of the moments from their expected values is considered by use of orthogonal transformations. These orthogonal transformations

yield alternative proofs of the distribution of χ^2 used in the goodness of fit without requiring the use of Stirling's approximation. The method of this paper is a generalization of the identification of the χ^2 used in approximating to the probability of obtaining exactly m successes in n trials with constant probability with the square of a standardized normal deviate derived from the consideration of n samples drawn from a population where the variable can take two values, 0 and 1. T. S. RUSSELL, *Virginia Polytechnic Institute.*

Lindley, D. V., "Statistical inference," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 30–76.

The paper is concerned with analysis of experiments and the point of view taken is that the purpose of experimentation is to enable one to decide between certain courses of action. Kolmogorov's axiomatic theory of probability is used and Wald's formulation of the statistical decision problem is adopted, but not in full generality. Two simplifying assumptions are made, namely, 1) decisions on experimentation have been made and the only decisions remaining to be made are terminal decisions; 2) both the class of probability distributions and of decisions are finite. Under these simplifying assumptions it is established that there exists a class of decision functions which is, in some sense, optimum. The concept of *minimum unlikelihood* is introduced and with it is constructed the optimum class of decision functions. Consequences of previous results are discussed as well as what must be specified in order that meaning be given to the phrase "the best decision function." Applications of the minimum unlikelihood method to some common statistical problems are given. Of particular interest is the result that $\bar{x}$ is the minimum unlikelihood estimator of the mean of a normal population for a wide class of weight functions. A discussion of the paper is included. F. S. MCFEELY, *Virginia Polytechnic Institute.*

Loevinger, J., Gleser, G. C., and DuBois, P. H., "Maximizing the discriminating power of a multiple-score test," *Psychometrika*, 18 (1953), 309–17.

Starting with a heterogeneous pool of items one frequently encounters the problem of constructing subtests in a manner that maximizes the correlation of items within subtests and minimizes the correlation between subtests. A method which seeks to maximize the ratio of inter-item covariance to total variance within each subtest is developed for this purpose. This ratio is defined to be the saturation of the subtest.

The nucleus for subtest 1 is formed from three items having high intercorrelations. In Cycle 1 those items in the pool which, when added to the nucleus, lower the saturation of the subtest are eliminated from further consideration in Cycle 1. Of the items that do not lower the saturation, that one is selected that maximizes the saturation. This process is continued at each stage with the augmented nucleus until the pool of items is exhausted. The construction of subtest 2 starts with a new nucleus of three items; Cycle 2 follows the same pattern as Cycle 1. Computationally several cycles can be carried out simultaneously.

Maximizing the saturation of a subtest drawn from a finite pool of items will not necessarily result in a battery having maximum discrimination power (minimum between subtest correlation). The conditions under which maximum discrimination is not achieved by the method are given. These conditions are not in general overrestrictive for practical use of the method. B. J. WINER, *University of North Carolina.*

Moran, P. A. P., "The random division of an interval—Part III," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 77–80.

This paper is a sequel to two previous papers on the subject by the same author. The distribution of the sum of the squares of the intervals into which the line is divided is further considered. The lower 5 per cent and 1 per cent points of the distribution can be found exactly up to $n=8$ and $n=9$ respectively (nine and ten intervals). Beyond about $n=20$ an approximation based on the variance ratio distribution is probably adequate. For values of n between 9 and 20 no workable formula has been reached. FRANKLIN S. MCFEELY, *Virginia Polytechnic Institute.*

Nath, Pran, "O. C. curve simplified," *Sankhya*, 13 (1953), 35–38.

It was first pointed out by Barnard that most O. C. Curves for fraction defectives were sufficiently well represented by a straight line, if drawn on logarithmic probability paper using the logarithmic scale for p. In this paper the author indicates that in his experience Barnard's method has not been satisfactory and he presents a method using "harmonic probability paper." Harmonic probability paper

is described and an illustration of its use is given. The author asserts that more satisfactory results are obtained using his method but gives no mathematical explanation. LEO LYNCH, *Virginia Polytechnic Institute.*

Ottestad, Per, "On the analysis of variance of percentage fractions," *Skandinavisk Aktuarietidskrift*, 35 (1952), 152–59.

This paper discusses weighting methods for the analysis of variance of percentages based on possibly unequal numbers of trials. Random variation of the binomial parameter and of the numbers of trials are taken into account. After some general comments on the meanings and limitations of the process, the author develops a regression method for obtaining the weights. Particular emphasis is given to the case in which number of trials is a random variable independent of the binomial parameter but dependent on classification. In this case he suggests that the weights should be found by estimating the (linear) regression of class variance for percentage as a function of the expected value of the reciprocal of sample size; the weights are then taken as the reciprocals of the appropriate values of the estimated regression function. An example taken from Cochran's article on the same topic in the *Journal of the American Statistical Association* 38 (1943), pp. 287–301, is worked out in detail. Finally the method is extended to cover the case of multiple classifications. SEYMOUR SUDMAN, *University of Chicago.*

Peto, S., "A dose response equation for the invasion of micro-organisms," *Biometrics*, 9 (1953), 320–35.

Where a dose can be represented as an integral number of units (say invading micro-organisms) and response is a yes-or-no event such as death, a one parameter model can be offered—viz. probability of survival equals e^{-np} where p is the unknown parameter. This can be offered as an approximation to $(1-p)^n$ where p is small and $(1-p)^n$ can be regarded as the probability of surviving the (independent) onslaughts of n organisms each having probability p of killing the host. Estimation by maximum likelihood is illustrated and tables to facilitate the (iterative) solution are given. The model is compared with the probit model and it is concluded that it is a suitable alternative for many problems. Efficient choice of dosage is shown to mean concentrating the doses in the 10%–35% survivors range. LINCOLN MOSES, *Stanford University.*

Reid, A. T., "On stochastic processes in biology," *Biometrics*, 9 (1953), 275–89.

There are many diverse fields of enquiry in the domain of biology where processes or mechanisms can best be considered in probabilistic terms. Examples arise in genetics, epidemiology, spread of rumors, organization and function of the nervous system, migration of organisms, population growth, experimental carcinogenesis. This paper considers illustrative problems which have been dealt with as stochastic processes. Unsolved problems are pointed out. The bibliography covers a large amount of recent work in this field. LINCOLN MOSES, *Stanford University.*

Rosenbaum, S., "Tables for a Nonparametric Test of Location," *Annals of Mathematical Statistics*, 25 (1954), 146–50.

To test whether two samples of n points and m points come from the same population, one counts the number of points, s, in one sample which lie outside an extreme value of the other sample.

Tables are constructed to show the probability is less than 1% (or 5%) that s or more points of a sample of size m ($\leqq 50$) lies outside an end point of a sample of size n ($\leqq 50$), provided the samples have been drawn randomly from the same population irrespective of its distribution. The author used some formulas given earlier by S. S. Wilks in a paper on tolerance limits. A. E. SARHAN, *University of North Carolina.*

Roy, P. M., "A note on the unreduced balanced incomplete block designs," *Sankhya*, 13 (1953), 11–16.

It is stated that an arrangement of vr units of v varieties, r units of each variety, in b blocks of size k ($k<v$) is known as a "Balanced Incomplete Block Design" if (i) a variety appears only once in a block and if (ii) pairs of varieties each appear in λ blocks. It is necessary that $bk=vr$, $\lambda(v-1)=r(k-1)$, and $b\geqq v$. A reduced form occurs when b, r, λ have no common factor. The object of the paper was to investigate (i) what are the unreduced designs, (ii) their connection with finite geometries, (iii) their connection with theorems of the method of differences, and (iv) whether they are capable of presentation in resolvable and affine forms suitable for the recovery of inter-block information. The only possibly resolvable forms are shown to be those designs where $v=2(t+1)$, $b=(2t+1)(t+1)$,

$r=2t+1$, $k=2$, $\lambda=1$. R. A. BRADLEY, *Virginia Polytechnic Institute.*

Sargan, J. D., "An approximate treatment of the properties of the correlogram and periodogram," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 140–52.

It is shown that the correlogram of an auto-regressive series can be regarded as though it were derived from an auto-regressive equation of double the order of the equation generating the original series. Some properties of the periodograms of time series generated by linear stochastic equations are investigated, and the results are applied to the study of Beveridge's wheat price index. G. L. EDGETT, *Virginia Polytechnic Institute.*

Sundrum, R. M., "The Power of Wilcoxon's 2-Sample Test," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 246–54.

Mann and Whitney's form of Wilcoxon's 2-sample test is described, and the variance of the statistic U under the null hypothesis is given. The upper bound of the variance of U is derived, and a population is constructed in which this upper bound is attained. Using the assumption that the statistic U is normally distributed under both the null and the one sided alternative, the power of the test is found for both normally distributed and uniformly distributed variates. The power of the test under the normal case is compared with the power of Student's t, and the power under the rectangular alternative is compared with the power of a test based on differences of mid-ranges and a test based on differences of sample means. H. C. SWEENY, *Virginia Polytechnic Institute.*

BOOK REVIEWS

Design for Decision. *Irwin D. J. Bross.* New York: The Macmillan Co., 1953. Pp. viii, 276. $4.25.

D. V. Lindley, *University of Cambridge*

"How do YOU make decisions? Are you unwittingly letting emotion conceal essential facts? Are you sure of making the right interpretations of facts and figures?" "How to make decisions that PAY. You'll learn new, more effective techniques for reaching the *best* decisions on questions of all kinds in—*Design for Decision.*" "The methods pay. Today a large University is paying a Ph.D. on its research staff more than its football coach—because of his decision-making skill." "Be the man in demand—get a copy of . . . ." And so on, and so on. The quotations are taken from the book-jacket of, or advertisements for, the book under review. It is a pity that in making a serious attempt to write popular science Bross should have been so ill-served by the blurb writers, and one can only query whether they weren't paid more than Bross for producing this rubbish.

In fact Bross has tried to give an account, as far as possible in non-mathematical language, of some of the ideas of modern statistical decision theory and statistical methods. Presumably the book is intended for the intelligent layman who wishes to have some idea of what these much-maligned statisticians do, and if so it succeeds reasonably well in giving a general impression, though on points of detail it falls very wide of the mark. The style is lively and, as far as I can judge, pure American. The blurb is wrong again in describing it as plain English. I found it most enjoyable. "[An] outline of history, from Ooze to Oak Ridge," "Statistical Inference. How to be a Great Detective in one easy lesson" are two delightful examples. Despite this style the author never makes the mistakes of the blurb writer.

The book falls naturally into two parts. The first begins with a brief account of the central role decision-making plays in man's existence, discusses the nature of probability, and the need for a value system in making decisions, and concludes with suggested rules for making decisions and some examples of their use. The second part contains an account of some modern statistical ideas, including a detailed account of the basic ideas involved in testing hypotheses and a briefer mention of other statistical tools. A supplement gives suggestions for further reading. Probability to Bross is all embracing; that is, we can speak of probabilities of hypotheses and use Bayes's theorem. (Minimax, as understood by him, only includes maximization over a restricted class of a priori distributions, namely the reasonable ones.) Value systems are essential and pragmatism is the philosophy of life. When decision rules are introduced they are not presented in the usual way with, on the one hand, the "states of the world" and, on the other, the possible decisions, in the way they usually occur in statistical problems, but with the decisions contrasted with the possible outcomes. This would make the appli-

cation of the rules to statistical inference difficult if it were not for the fact that Bross never seems to realize that statistical decision theory has anything to do with, for example, testing the mean of a normal population when the standard deviation is known! In fact the second half of the book is independent of the first half, apart from the discussion on probability. This is quite the most astonishing thing about the book. Whilst on p. 83 we are told that "the methods proposed by R. A. Fisher to replace Bayes rule also make assumptions about the [prior probabilities] . . . and the substitutes for Bayes rule actually represent special cases of the rule," we are later (Chapter 13) presented with a Fisherian argument for the test of normal mean based on significance levels without a mention of Bayes's theorem! Similar remarks apply to the value system, except for some observations on a Simple Value System (p. 101) which are not explored. It is amazing that he should not realize that a logical pursuit of the decision function method (either with or without prior probabilities) leads to a rejection of the significance level approach: in general if weights are assigned without consideration of the sample size, as they usually would be, then the significance level (i.e. the probability of error) will vary with this size. Consequently the second half of the book is not merely independent of, but is in contradiction with, the first part. There is even more to it than that, for some of the paradoxes of standard statistical methods are quoted without it being realized that they are resolved by Wald's ideas. Thus (p. 209) "One crusader against the myth of normality has a standard offer of $100.00 for any collection of data with over one thousand observations which will meet the standard statistical tests of normality. So far as I know he has had no takers." Decision theory—balancing one hypothesis against another—resolves this criticism of statistical tests.

Nevertheless the second part of the book does contain a lively, reasonably accurate, and always interesting account of some statistical ideas. This is possible because, so far as one can see, practicing statisticians, that is people who handle data, not the mathematical manipulators of uninterpreted symbols, do not use decision theory ideas. Significance levels are still used despite Wald. To discuss why this is so would take us too far away from the book under review. Any statistician wishing to recommend this book to his friends should therefore warn them to take a good pinch of salt with the first part. Even the second half is not free from blemish for there is one grand howler. The test chosen by Bross as an example is that for the mean of a normal population, standard deviation known. The description is lucid and it is a great pity that it is spoiled by the use of the sum of squares as the test criterion and reference to the χ^2-table for its significance! (p. 229). This appears to have arisen through a misunderstanding of the meaning of the term "sufficient statistic."

There are many other things one might comment on, for this is a stimulating book. Several remarks are sheer nonsense: "Our standard for bias is based on the rule that whenever an overwhelming majority of observers are

in agreement they are, ipso facto, right" (p. 152). Others are very sensible—those on the importance of a value system, for example. Other remarks are contradictory: "if this outcome or result of the decision process is agreeable to *me*, then the decision may be adjudged satisfactory" (p. 20), but " . . . a definite and *objective* way to progress from data to inference . . . all arrive at the same conclusion provided they start from the same data" (p. 221). (In both cases my italics.) What does, however, annoy me is Bross's assurance. He states everything as though there were no doubt at all. Pragmatism is right. Probabilities of hypotheses are all right. He never gives a hint that other people might hold different views and yet be sound. The world is essentially simple to him: his blacks are jet, his whites are bright and there are no greys. I envy him. But "I beseech you, in the bowels of Christ, think it possible, you may be mistaken."

Here then is a book which gives overall a good account of statistical ideas for the layman, a less satisfactory account of decision theory, but which is always entertaining. It contains a magnificent misprint. In connection with the Renaissance, the renewed interest in Greek ideas, and the dawn of the experimental method, we read: "The same era that witnessed the rediscovery of Reason also saw the birth of the successor to Reason—Silence."

The Design and Analysis of Experiment. *M. H. Quenouille.* New York: Hafner Publishing Co., 1953. Pp. xiii, 356. $7.50.

BERNARD OSTLE, *Montana State College*

THE appearance of four books on experimental design in the last few years is evidence that authors have become aware of a gap which has existed in the literature for too long. The four books are: *Experimental Designs* by Cochran and Cox, *Design and Analysis of Experiments* by Kempthorne *Analysis and Design of Experiments* by Mann, and the volume reviewed here. The four authors had different goals. Mann's text is a purely mathematical approach to general linear hypothesis theory. Cochran and Cox give a catalog of useful designs, accompanied by considerable explanatory material which makes their book most helpful to research workers. Kempthorne's work is conceived on a larger scale: he successfully attempts a logical development of designs (mainly factorials) and design principles, and thus has produced a book of an advanced nature more suitable for graduate study than for research use. Quenouille's book is most nearly akin to that by Cochran and Cox, but gives greater emphasis to groups of experiments and long-term policy.

The *Design and Analysis of Experiment* has been divided into four sections which, with the topics included in each, are: (A) Elementary Principles and Designs: (1) The design and analysis of experiments, (2) Randomised blocks and Latin squares, (3) Factorial and split-plot designs; (B) Incomplete Block Designs: (1) Factorial designs involving factors at two or three levels,

(2) Complex factorial designs, (3) Incomplete block designs for a single set of treatments; (C) Long-Term Policy: (1) Long-term experiments, (2) Planning of groups of experiments, (3) Combinations of experimental results; (D) Experimental Complications: (1) Special designs and analyses, (2) Missing observations, (3) Scaling of observations.

In the preface, the author states: "This book is aimed at those wishing to acquire a working knowledge of experimental design and an understanding of the principles governing it." Unfortunately, the explanations are often too brief and too technical to be of great value to research workers not well-versed in statistical theory and principles. In particular, this is not a book for persons lacking a solid background in analysis of variance. The constant stress on the assumptions necessary for a valid use of the various designs is, however, an excellent feature. Too often this important part of experimental planning is omitted from classroom lectures and text book chapters.

In section B, Quenouille describes the main types of confounding. The use of partial confounding is discouraged because of complexities of computation. While it is true that partial confounding does add to the time involved in calculation, this extra effort is frequently worthwhile. This apparent condemnation of partial confounding is characteristic of Quenouille's tendency to use sweeping statements which may lead to the non-critical reader to lose the benefits of certain advanced technqiues.

The most interesting material in this book is contained in Sections C and D, especially under the heading Long-Term Policy. The discussion on planning of groups of experiments and combining experimental results is excellent. On the other hand, the reviewer wonders why "missing observations" were singled out for special attention. This subject could easily have been handled as a necessary consideration when discussing each particular design. Also, rotation experiments surely deserve more attention than the four pages devoted to them.

In summary, much useful information is contained in the book. However, the tendency to be *brief* and *technical*, accompanied by an occasional carelessness in writing (due, no doubt, to attempting to complete the work within too short a period of time), has detracted from the value of the book. It is not to be read quickly. Nevertheless, experienced research workers in the agricultural and biological sciences will find it a good reference text if read critically.

Sample Survey Methods and Theory. Vol. I. Methods and Applications; Vol. II. Theory. *Morris H. Hansen, William N. Hurwitz and William G. Madow.* New York: John Wiley & Sons, Inc., 1953. Pp. xxii, 638; xiii, 332. $8.00; $7.00.

TORE DALENIUS, *Stockholm*

Introduction. The campaigns conducted during the first quarter of this century for an increased use of "the representative method," supported

by the International Statistical Institute among others, showed special success in the latter part of the 1930's. The lead in the new era of sampling was taken by India and the United Kingdom, where the development naturally was knitted to improvements in the field of agricultural statistics, and by the United States, where a great portion of the development was knitted to improvements of methods for measuring socio-economic phenomena.

The developments of 1939 accelerated this trend. In the U. S., the Bureau of the Census took the lead. The Census Bureau introduced sampling methods into the 1940 census to an extent not previously seen, and transferred many sample surveys carried out by means of non-probability methods to a probability basis. This development was responsible for the creation of a large and competent "sampling staff" within the Census Bureau. Among the members of this staff were Morris H. Hansen, William N. Hurwitz and William G. Madow, authors of the two volume work *Sample Survey Methods and Theory*, published as one of the Wiley Publications in Statistics.

A large amount of the work carried out by Hansen-Hurwitz-Madow and their colleagues necessarily meant application of already available theory to actual survey operations. But a considerable portion of the activity was devoted to the development of new theory and new methods.

Portions of the results thus achieved have been presented earlier; examples are the book *A Chapter in Population Sampling*, the almost classical 1943 paper entitled "On the Theory of Sampling from Finite Populations" and the 1949 paper "On the Determination of Optimum Probabilities in Sampling."

Objects of the book. Sample Survey Methods and Theory represents "an attempt to give a comprehensive presentation of both sampling theory and practice." The book as a whole, as well as each one of the two separate volumes, is designed as a textbook; it should, moreover, serve as a manual for the investigator engaged in the design of sample surveys. Finally, parts of the book are intended for "the user of the results of surveys who wishes to know the circumstances under which he may place confidence in information based on samples."

Broad summary of content. As indicated by the subtitles of the two volumes, Volume I is devoted to applications (it is labelled, by Wiley, "applied statistics"); Volume II is devoted to theory (labelled, by Wiley, "mathematical statistics").

Volume I may be looked upon as made up by three parts; the introduction and chapters 1–3 make up the first part, chapters 4–11 the second part, and chapter 12 the third part.

The introduction and chapters 1–3 address themselves to the *consumer* of survey results rather than to the producer. The language is "nonmathematical"; this does not mean a complete lack of formulas and symbols but a frequent use of easily-grasped illustrations. In addition to presenting the usual sampling principles, the first part presents the philosophy of the use of measurable sample survey designs.

Chapters 4–11 present a detailed account of *methods* for sampling from finite populations. Most results are given with references to proofs in Volume II. The following list of chapter headings gives an idea of the contents: 4, Simple random sampling; 5, Stratified random sampling; 6, Simple one- or two-stage cluster sampling; 7, Stratified single- or multi-stage cluster sampling; 8, Control of variation in size of cluster in estimating totals, averages, or ratios; 9, Multi-stage sampling with large primary sampling units; 10, Estimating variances; 11, Regression estimates, double sampling, sampling for time series, systematic sampling, and other sampling methods.

Chapter 12, "Case Studies," constitutes the third part. In the first half of this 110 page chapter, three sample surveys from the practice of the Census Bureau are presented in detail in a way which demonstrates how the many methods presented in chapters 4–11 are integrated into a survey *design*. The rest of chapter 12 deals with two studies of variances and co-variances and the use of quality control methods in the office processing of the 1950 Censuses of Population, Housing, and Agriculture.

Likewise, Volume II may be looked upon as made up of three parts; chapters 1–3 make up the first part, chapters 4–11 the second part and chapter 12 the third part. In chapter 1, the fundamental definitions used in Volume I are summarized; by this device, used throughout Volume II, this volume is made self-contained; it is possible to read Volume II before Volume I, or even to read only Volume II. Chapters 2–3 present the fundamental theorems on probability, expected values and variances necessary for mastering the proofs given in the following chapters.

Chapters 4–11 are mainly devoted to proofs of the results quoted in Volume I. The presentation runs, chapter by chapter, parallel in the two volumes. However, in addition to the empirical results presented in Volume I, there are new ones scattered in these chapters.

Chapter 12, finally, presents a theory of response errors; the chapter is a revision of a paper published in this *Journal* in 1951.

An effort to evaluate the book. I am of the opinion that a review, in order to be comprehensive, should end up with an effort to evaluate the book against the background of the objectives set forth by the authors.

The book *is* a "comprehensive presentation of both sampling theory and practice." As to *theory* (for sampling from finite populations), I am at a loss to find anything of importance that has been left out; as to *practice*, it is true that most applications are selected from the work within the Census Bureau but there seems to be no important type of finite population entirely left out.

I have not had experience with this book as a textbook; I have had, however, the opportunity of attending a course, in Washington in 1951, which was based on notes having a story in common with this book. From this experience, I conclude that the book will be found to be an excellent textbook.

There is a specific feature of the book which deserves to be mentioned. By splitting the book into two volumes, one "applied" and one "theoretical,"

the authors have been able to present the *solution* of a design problem in different ways in the two volumes. In Volume I, the authors stress just as much the region (interval or whatever it may be) within which the exact mathematical solution is to be found, as the exact solution itself; thus, the authors stress that "the optimum is broad" when discussing, for example, optimum allocation in stratified sampling and optimum size of cluster. Solutions of this kind are sorely needed in actual work where one almost always has to design without exact information as to the size of important "design parameters" (such as variances). In volume II, on the other hand, emphasis is on the exact solutions. Volume II thus teaches the technique to use on one's own problems.

Space does not permit a detailed discussion of other valuable aspects of this book. I only want to mention that I welcome especially the thorough analysis of survey costs and the construction of cost functions. There are, in official survey reports from all over the world, many examples of cost functions; but these examples are often difficult to interpret and use as long as there are only, at the very best, indications of what kind of costs are taken care of by the different components in the cost functions.

In summary, this is a great book, which will be indispensable to every person, statistician or not, who comes close to sample surveys. Of course, in a book of nearly 1,000 pages, one can find opportunities to criticize some points. Most of them are too minor to be dealt with at length (e.g., the definition of simple random sampling, chapter 4, does not seem to fit the one in chapter 5; I would prefer to see a distinction between a "ratio estimate" and an "estimate of a ratio," and so on). But it is perhaps justifiable to say that chapter 11 is somewhat displaced and too "mixed." Personally, I would rather see one separate chapter devoted to systematic sampling, possibly placed before the present chapter 6. The difference and regression estimates could be discussed in exactly the same way as are the ratio estimates (i.e., in conjunction with the specific sampling systems, as means of improving precision over and above that of simple estimates such as the sample mean). [Note: Since chapters numbers and context are parallel in the two volumes, the foregoing reference to chapters are for both volumes.]

The language in the book is, even for a foreigner, easy; one soon gets used to words like "rel-variance," "epsem," etc. But the symbols are part of the language. The lack of a *standard* in this area is in itself regrettable and should, I think, be taken care of. As a result of this lack, Wiley has published in 1953 two excellent books in sampling, this and Cochran's, which use rather different symbols.

I hate to finish my review of this excellent book by being critical, so I repeat: *Sampling Survey Methods and Theory* is a great book, which will be indispensable to every person, statistician or not, who comes close to sample surveys.

Elementary Statistical Analysis. *Harry P. Hartkemeier.* Dubuque, Iowa: Wm. C. Brown Co., 1952. Pp. xxii, 484. $6.00. Paper.

ACHESON J. DUNCAN, *The Johns Hopkins University*

THIS book by Professor Hartkemeier is the text that is used in the freshman course in elementary statistics at the University of Missouri at which he is Director of the Statistics Laboratory. It assumes no prerequisite of college mathematics and is slanted toward the student of business. In the author's own words, it "has been written for the person who likes to have directions for the immediate practical application of the elementary statistical techniques to sample data without waiting until all statistical techniques and the mathematical theory underlying them have been explained in detail."

In his twenty odd years of teaching statistics Professor Hartkemeier has developed many novel ideas as to how a beginning course in the subject should be taught and has as a consequence written a book that is strikingly different from other statistics texts. This is true with respect to both format and contents. The pages are full size reproductions of typewritten copy, the cover is heavy paper and the whole is fastened together with plastic screws that allow withdrawal of pages at will. This loose leaf character of the text permits the author to include special problem forms and work sheets that upon completion may be extracted by the student for submission to the instructor. It also permits extraction of tables at time of examination without running the danger of recourse to forbidden sections of the text. Illustrations, tables, and problem forms are placed at the ends of the chapters thereby permitting ready reference.

The book is concerned primarily with tests of significance for small samples. After an "introductory" chapter in which the reader is given the elements of frequency distributions, time series analysis, index numbers and correlation (all in 35 pages), the author settles down to a detailed discussion of the computation of square roots and use of mathematical tables, the computation and use of the arithmetic mean and standard deviation and other types of averages and representative values, tests of significance for means and standard deviations and comparisons of two means and standard deviations, contingency tables and chi-square tests, and a discussion of analysis of variance that includes problems related to both single and two way classifications, unequal numbers in different cells, Latin squares, and data with several sources of random error. There is also a chapter on statistical quality control in manufacturing operations, and a chapter on computing procedures and machines, drawn heavily from the author's book *Punch-Card Methods* and illustrated copiously with pictures (supplemented with directions for use) of many of the modern computing machines, including even the Monroe Octal Adding-Calculator. All this material is presented in a very readable style that the reviewer believes beginning students will like. At the end of each chapter there are numerous problems and considerable problem data.

In writing a beginning book that attempts to explain the principles of statistical theory without the use of mathematics, the greatest difficulty is to prevent the argument from becoming inexact and perhaps incorrect. Although Professor Hartkemeier does a splendid job on the whole of explaining difficult material, he has not succeeded entirely in avoiding misleading statements and errors. The principal instances of this kind noted by the reviewer are as follows:

(1) Charts of frequency curves in the book have a vertical scale marked "number of" or "relative number of" cases, whereas in fact it is the area under the curve that is a measure of the "relative number of" cases and not the ordinate of the curve.

(2) The text suggests that the mean has little use in a highly skewed distribution. There are cases, however, in which it may be the "best" average for certain purposes. If we know the mean family income, for example, and the number of families in a given community, we can compute the community income. We cannot do this with the mode or the median.

(3) The text tends to give a false conception regarding the character and use of the t-distribution. The impression is given that the ratio of a variable to its standard error follows the t-distribution if the sample from which the variable is calculated is small. Thus, on p. 372, the ratio of a mean of a sample of 25 to a *known* standard error is treated as a t variable. Actually it is a normal variable. Again, on p. 373, the ratio (σ sample $-\sigma$ universe)/known standard error of σ is treated as if it had the t-distribution (apparently because the sample size is 25) whereas, it actually is distributed as χ or more precisely as $\sqrt{2}(\chi-\sqrt{N})$.

The facts about the t-distribution are as follows: If z is a normally distributed variable with zero mean and unit standard deviation, if u is a variable that follows the χ^2 distribution with n degrees of freedom, and if z and u are independently distributed, then the ratio $z/\sqrt{u/n}$ has the t distribution with n degrees of freedom. The t distribution approaches the normal distribution as the degrees of freedom become infinitely large.

(4) The text fails to point out (p. 334) that the use of the ordinary χ^2 table of probabilities in χ^2-tests of frequencies involves an approximation. This is essentially the same in character as the approximation of a binomial probability by an area under a normal curve. It is the reason for the χ^2 correction for continuity in a 2×2 contingency table when samples are not large.

(5) The title of Chapter 9, "Statistical Methods Necessary for Quality Control in Manufacturing Operations," is misleading. A better one would have been "The Author's Ideas on Some Quality Control Procedures." Although Professor Hartkemeier refers to several books on quality control, including one by the reviewer, he draws little from them. Instead he proceeds to describe methods that differ widely from the standard procedures. Thus, his $\overline{X}$-chart uses samples of 25, not the customary samples of 4 or 5.

The central line on the chart is apparently based on specifications, in the manner of a modified $\overline{X}$-chart. The lower limits on the chart are 2.5% and 0.5% probability limits incorrectly based on the t-distribution, not the ordinary 3 σ limit. The upper limits are statistical limits but are confusedly interpreted as being related to some specification based on the desire not to produce too good a product. The discussion emphasizes the use of the $\overline{X}$-chart to maintain a constant allowable fraction defective, but says little about the use of the chart in discovering assignable causes (the principal use of the control chart).

The chart that is suggested for controlling variability is the little-used standard deviation chart. Again it is related to the attempt to maintain a constant fraction defective rather than to the attempt to detect assignable causes. Strangely (and incorrectly), limits are again set by a t factor. No mention is made of the range chart, although this is used in industry many more times than the standard deviation chart. The statement made earlier in the text (pp. 117, 212) that the range is not used much by statisticians is very much in error in the industrial field.

In any course in which this book is used, the reviewer would strongly urge the omission of this chapter on statistical quality control. Fortunately, such an omission can be made without difficulty.

(6) The text does not state the basic assumptions of analysis of variance (additivity of main effects and normality, independence, and uniform variability of the random variable). This omission is somewhat unfortunate in that the text is free and easy in its illustrations of the uses of analysis of variance. It is applied to percentage data (pp. 441, 451) where the assumption of uniform variability is probably violated; and it is applied to sales data (p. 443) where the assumption of additivity of main effects is questionable. It may be admitted that the F-test is "robust" and may not be seriously affected by such deviations from strictly valid procedures. The use of such illustrations, however, coupled with failure to state the assumptions underlying strictly valid procedures may lead the reader to believe that there are no limitations to the application of analysis of variance techniques.

More serious than these deviations from the basic assumptions of analysis of variance is the tendency to play around with the data until some significant conclusion is reached with little regard to the effect of this procedure upon the final level of significance. Thus, t-tests are run after an F-test shows non-significance (p. 398), and data are reclassified (p. 399) and tested again after a first F-test shows non-significance.

The use of charts to show interactions is an excellent device, but to interpret an interaction as the "crossing" of the movements at various levels (p. 409) is not accurate. The criterion of interaction is non-similarity of movement, not crossing of the interaction graphs.

The computation of a row sum of squares (p. 394) when there are actually not rows but merely an equal number of cases in each class is very confusing and should be omitted.

These criticisms should not deter an instructor well trained in statistics from using the text for a beginning course for, as noted above, it will probably be liked by the students and this is important. A greater drawback in the eyes of the reviewer is the little attention given in the book to confidence intervals, errors of the second kind, and correlation.

Mathematics and Statistics for Economists. *Gerhard Tintner.* New York: Rinehart and Company, Inc., 1953. Pp. xiv, 363. $6.50.

G. Baley Price, *University of Kansas*

The standard mathematics curriculum in American colleges and universities is one which has grown up in connection with the physical sciences—one which has been designed to support the study of chemistry, physics, and the engineering sciences. The traditional sequence of courses—college algebra, trigonometry, analytic geometry, calculus, and differential equations—is badly out of date because it has undergone no fundamental revision in fifty years, and perhaps a hundred.

There is abundant evidence that major changes are in progress. The Mathematical Association of America and the National Council of Teachers of Mathematics sponsored a Conference on Teacher Training at the University of Wisconsin in the summer of 1952. As an outgrowth of this conference and of the activities of the National Research Council's Committee on the Regional Development of Mathematics, two committees have been appointed to study the revision of the undergraduate curriculum. One is a joint committee of the MAA and the NCTM under the chairmanship of Dr. C. V. Newsom, and the other is a committee of the MAA under the chairmanship of Professor W. L. Duren, Jr. The National Science Foundation sponsored a Summer Conference on Collegiate Mathematics at the University of Colorado in the summer of 1953, and it will sponsor two similar conferences, at the University of Oregon and the University of North Carolina, in the summer of 1954. The National Science Foundation will also sponsor a conference for high school teachers of mathematics in the summer of 1954. All of these activities are designed to modernize the mathematics curriculum and its teachers.

Professor Tintner's book on *Mathematics and Statistics for Economists* must be considered further evidence of the change that is taking place in mathematics, especially in its relation to the social and biological sciences. This book was written to teach mathematics and statistics to economists, especially to future econometricians. It was written for students who know economics, and who have some knowledge of algebra and trigonometry.

The book is divided into three parts. Part I covers pages 3 to 65. The chapter headings in this part are: functions and graphs; linear equations in one unknown; systems of linear equations; quadratic equations in one unknown; logarithms; progressions; determinants; and linear difference equations with constant coefficients. It might be supposed from these chapter headings

that Part I is a brief college algebra, but this is not the case: it is entitled *Some Applications of Elementary Mathematics to Economics.* In the first place, the treatment of algebra is far less extensive than the chapter titles would suggest. The chapter on the quadratic equation in one unknown contains a half page of discussion, in which the formula for the roots of the quadratic is stated, and a half page of exercises. The chapter on logarithms contains no treatment of the properties of logarithms and their applications to computation. Instead, a knowledge of logarithms is assumed, and applications are made to two problems in economics. The chapter on determinants is marked as one that is not needed for the remainder of the book and can be omitted. On the other hand, a chapter on linear difference equations is included, and it is not marked as one that can be omitted. Thus the algebra content of Part I is far less than that of the typical course in college algebra. In the second place, Part I contains a large amount of economics. Indeed, it contains a treatment of the following topics from economics: linear programming; linear-supply functions; linear-demand functions; market equilibrium; market equilibrium for several commodities; imputation; quadratic demand and supply curves; Pareto distribution of income; demand curves with constant elasticity; growth of enterprise; population theory of Malthus; and compound interest.

Part II, entitled *Calculus*, covers pages 69 to 190 and contains a fairly extensive treatment of differential calculus. One chapter of 13 pages is devoted to a treatment of integral calculus. This chapter states the fundamental theorem of calculus, but it does not suggest a proof. The extent of the treatment of calculus is well indicated by the chapter headings in Part II: functions, limits, and derivatives; rules of differentiation; derivatives of logarithmic and exponential functions; economic applications of the derivative; additional applications of derivatives; higher derivatives; maxima and minima in one variable, inflection points; derivatives of functions of several variables; homogeneity; higher partial derivatives and applications; elements of integration. It may be remarked that the treatment of calculus given here is far less extensive than that contained in the standard calculus courses normally given in the freshman and sophomore years to physical science students. Part II also treats the following topics in economics: demand functions and total revenue functions; total and average-cost functions; marginal cost; marginal revenue; elasticity; elasticity of demand; marginal revenue and elasticity of demand; increasing and decreasing marginal costs; monopoly; average and marginal cost; marginal productivity; partial elasticities of demand; joint production; utility theory; production under free competition; marginal cost, total cost, average cost; and consumer's surplus.

Part III, entitled *Probability and Statistics*, fills pages 193 to 309 and contains a brief but significant treatment of probability and statistics. The chapter headings are the following: probability; random variables; moments;

binomial and normal distributions; elements of sampling; tests of hypotheses; fitting of distributions; regression and correlation; index numbers; and a postscript which contains suggestions for further reading. In contrast with the first two parts of the book, economic theory is conspicuous by its absence from Part III. To be sure, the fitting of demand and supply curves is treated on pages 292 to 297, but in general this part of the book appears to be a rather straightforward treatment of statistics.

Pages 311 to 340 contain answers to the odd-numbered exercises. The next section of the book consists of six tables as follows: four place common logarithms (pp. 341–342); natural trigonometric functions (pp. 343–346); four place natural logarithms (pp. 347–348); area of the normal probability curve (p. 349); Student's t-distribution (p. 350); and the χ^2 probability scale (p. 351). The book ends with an index of names (p. 355); an index of mathematical and statistical terms (pp. 357–360); and an index of economic terms (pp. 361–363).

Many departments of mathematics now recognize a reponsibility to teach mathematics for the social and biological sciences as well as for the physical sciences. Professor Tintner's book emphasizes a number of the problems that face these departments in their efforts to discharge their new responsibilities. The typical instructor in mathematics will feel that he is not competent to teach *Mathematics and Statistics for Economists* because his knowledge of economics and statistics is inadequate. A major problem in the introduction of a new undergraduate curriculum will be the training of the present mathematics staffs to teach the new curriculum. The fact that Professor Tintner holds the unusual title of Professor of Economics, Mathematics, and Statistics emphasizes that he is not a typical staff member. A major problem concerns the organization and arrangement of a curriculum designed to serve the needs of all those fields which now make significant use of mathematics. Must departments of mathematics now offer separate courses for physicists and chemists, for engineers, for economists, for psychologists, for biologists, and so on? The small liberal arts colleges, in which so many of our scientists and other scholars originate, will find such an arrangement impossible because of their limited staffs and the small number of their students. Many university educators and administrators will oppose such specialized courses on a variety of grounds. But is it possible to devise a freshman course in mathematics that the entire university will find adequate and acceptable? The question is more easily asked than answered. Finally, Professor Tintner's book emphasizes that the needs of the economists will not be satisfied by a course in statistics. It appears significant that, as pointed out above, the concepts and principles of economics occur in Parts I and II rather than in connection with statistics in Part III.

Whatever the ultimate solution of the problems involved in teaching mathematics to economists, both the mathematicians and the economists are indebted to Professor Tintner. He has written what appears to be a teachable

textbook in a new field. In particular, he has provided an enormous collection of significant and vital exercises in economics which involve the elementary parts of mathematics. The extent of this collection of exercises is indicated by the fact that, as noted above, the answers to the odd-numbered exercises alone fill 30 pages of the book. One of the major problems that confronts the mathematicians in their efforts to revise their elementary curriculum is the collection of similar exercise material which will relate the old and new mathematics to the various fields and subjects where it finds application.

Hood, William C., and Koopmans, Tjalling C., editors. *Studies in Econometric Method.* Cowles Commission Monograph No. 14. New York: John Wiley and Sons, 1953. Pp. xix, 323. $5.50.

KENNETH J. ARROW, *Stanford University*

THIS collection of ten studies of problems in the estimation of simultaneous structural equations is indispensable for the modern econometrician. It gives a virtually complete picture of the present state of the subject and at the same time is eminently readable. Of the papers, three have been previously published, the remainder being specially written for this volume.

The present volume follows a pattern already set in earlier collective works published by the Cowles Commission; there is one long paper which sets forth systematically the basic ideas of the subject, while the remaining papers present more detailed expositions or further developments. In this case, the central paper in Chapter VI, by Koopmans and Hood, an admirably clear exposition of the model of linear simultaneous stochastic difference equations, the definition of exogenous, endogenous, and predetermined variables, the criteria for identification in this model, the maximum-likelihood method of estimation with particular reference to the limited-information case, and some statistical tests of the validity and the identifying power of *a priori* restrictions. The derivations are new and very much simplified from earlier versions, though perhaps one would hardly call them simple in an absolute sense. Though the article is not written in textbook form, it will form essential supplementary reading to a book such as Klein's, if it is desired to supply the student with a derivation of the basic estimation formulas.[1]

The first paper, by Jacob Marschak, is an excellent exposition of the role of statistical inference in economic policy and prediction. The concepts studied by Koopmans and Hood are here introduced in relation to the uses for which they are intended. The second paper, by Koopmans, is a nontechnical exposition of the concept and problems of identification; it is reproduced, with minor revisions, from a paper published in *Econometrica.* The careful discussion of various examples will be invaluable pedagogically. In the third paper, Herbert A. Simon seeks, on the basis of the concepts of exogenous and endogenous variables, to define the notion of causality in a

[1] L. R. Klein, *A Textbook of Econometrics,* Evanston and White Plains: Row, Peterson, and Company, 1953.

way which will meet positivist objections, such as those of Hume. He argues that the complete rejection of the concept of causality (as opposed to functional interrelationship), as in Russell's position (to which, however, Simon does not refer), does not correspond to the intuitive practice of scientists. An interesting discussion is then given of causal ordering of variables in a linear structure; the concept is closely related to that of identifiability.

The fourth and fifth papers, by Trygve Haavelmo, and by M. A. Girshick and Haavelmo, respectively, are now famous empirical applications of the simultaneous equations method to the consumption function and to the demand for food.

The seventh paper, by Herman Chernoff and Herman Rubin, shows, without proofs, how limited-information estimates may be used even when the conditions under which they were derived are not valid, in the sense that they give rise to consistent estimates. From the point of view of new knowledge, this is undoubtedly the most important paper of the volume. It is remarked that if predetermined variables in the system but not in the group of equations to be estimated are omitted, the estimates resulting will still be consistent if the variables omitted are not needed for identification. It is also shown that in many cases errors in the variables and non-linearities in the equations can be accommodated.

In the eighth paper, Stephen G. Allen studies, in an example, the loss of efficiency by omitting a predetermined variable in a particular equation to be estimated. In the ninth paper, Jean Bronfenbrenner (Crockett) examines the bias attributable to the use of the method of least squares in a two-equation model. Both papers are very illuminating in giving a more intuitive appreciation of the sense in which the simultaneous equations methods is optimal.

The last paper, by Herman Chernoff and Nathan Divinsky, is an extremely complete exposition of the computational methods used in various types of maximum-likelihood estimates. The practicing econometrician will make extensive use of this section.

This is a very useful collection of papers, which I can strongly recommend.

Stochastic Processes. *J. L. Doob.* New York: John Wiley and Sons, 1953. Pp. vii, 654. $10.00.

P. A. P. MORAN, *Australian National University, Canberra*

THE average statistician or worker in applied probability theory never has to deal with more than a finite number of random variables and thus never really needs to know much about measure theory. But the mathematician who wishes to found probability theory on a rigorous basis needs more elaborate theory. The strong law of large numbers (which is essentially an empirically unverifiable theorem) necessarily involves a theory of measure in a space with an enumerable infinity of dimensions while the theory of con-

tinuous random processes requires, for a fully rigorous foundation, a deep discussion of measure theory. In the present very remarkable book Professor Doob sets out a fully rigorous foundation for the theory of random processes both with discrete and with continuous parameters and in the course of doing this discusses a number of theorems of interest to those working on other parts of pure probability theory. The result is the most complete discussion yet published in book form of the foundations of the theory of processes.

The book opens with a chapter on probability theory which is openly and frankly (and in the opinion of the reviewer, rightly) equated with the theory of measure. The elementary ideas of random variables, families of variables, and modes of convergences for random variables are very carefully introduced and then follows the most important part of this chapter—a discussion of conditional probabilities. This is usually skirted around in textbooks since a rigorous treatment requires considerable care, and was first given by Kolmogoroff. Then, after some standard results on characteristic functions there are some very interesting new inequalities between the tails of a distribution and integrals involving its characteristic function.

In the second chapter the author introduces the idea of stochastic process and considers the classification of such processes in general. For a process in which "time" (here always represented by a variable on a linear set) is continuous we immediately get into difficulties in setting up a probability measure in the class of all realizations of such a process. These are overcome by insisting that the processes be "separable" in a certain sense introduced by the author. This is probably the most difficult part of the book to the average reader and requires a knowledge of measure theory beyond that of most analysts. A useful supplement at the end of the book attempts to bridge the gap. Next, the author introduces Gaussian processes and gives a discussion of the Markovian property which is rather more careful than is usual.

The next chapter on processes with mutually independent variables is concerned with classical results in the theory of probability, most of which will be more or less familiar to the reader. These are not discussed for their own sake but for the light they throw on random processes. The Borel-Cantelli lemma and similar results on series of variates and the law of large numbers are discussed with great care. Next we have a welcome account of infinitely divisible distributions and Lévy's general formula for their characteristic functions. The arithmetic of distributions is not, however, considered further than the needs of the theory of processes requires. A general account of this subject in English is much to be desired but remains to be written.

Chapter IV considers processes with mutually uncorrelated but not necessarily independent variables. This is thus more general and the interest lies in seeing how far we can get with the weaker assumption.

Chapter V on Markov chains with a discrete parameter deals with relatively familiar theory. The usual theory of a finite Markov chain with sta-

tionary transition probabilities, and the classification of the states, is given together with a short application to card mixing. Multiple Markov chains in which the transition probabilities depend not merely on the previous states but also on earlier states are reduced in the obvious way to ordinary Markov chains. This, however, is an illustration of the way in which the author deals only with the general theory of the subject (difficult as that is) without dealing with the analytical difficulties which arise as soon as we specialize the theory to some particular problem. Complex Markov chains (as Markov originally called them) are very awkward to deal with in practice. Next we have a long and rather difficult account of the generalization (mainly due to Doeblin) of the previous theory to general state spaces and the corresponding law of large numbers and a central limit theorem.

Next we have a chapter on Markov processes with a continuous parameter, firstly those with a finite number of states with the usual theory of forward and backward differential equations and then to a continuous state space and chains with an enumerably infinite set of states, the latter being nowadays of very great importance in such subjects as the theory of queueing. The Fokker-Planck and related equations are then considered at some length.

Chapter VII is a long (100 pages) and interesting chapter on what are now known as martingales. This word, which is due in this connection to J. Ville, is usually used in connection with the harness of horses and the rigging of ships but is also used, for some odd reason, for a gambling system in which the stake is increased by a factor of two at each trial. In probability theory a martingale is defined as a random process $\{x_t\}$ such that $E\{|x_t|\} < \infty$ and

$$x_{t_n} = E\{x_{t_{n+1}} | x_{t_1}, \cdots, x_{t_n}\}$$

with unit probability, whenever $t_1 < \cdots < t_{n+1}$, and n is an arbitrary positive integer. This watery looking object does not look at first sight as if it would have a very interesting theory, but the theorems and results which follow are most interesting and varied and introduce a new unity into a widely scattered series of problems. Much of the work described is due to the author.

He first considers applications to games of chance where the idea of martingale is used to provide a definition of a "fair" game. This leads to a new and interesting discussion of the effect on fair games of systems of optional stopping and sampling. Next follow some new inequalities for expectations in martingales and a series of convergence theorems. The general theory is then applied to sums of independent variables, to the strong law of large numbers, the theory of derivatives, and the relationship to the likelihood ratio and sequential analysis pointed out. The corresponding theory of martingales with a continuous parameter is then developed at some length together with some remarks on the application to Poisson and Brownian processes.

In the eighth chapter processes with independent increments, which form a logical prelude to the study of stationary processes, are considered.

Examples of these are provided by the elementary Poisson process and the Brownian movement. The centering of the general process of this type and the classical theory of the character of the distribution function (due to its infinite divisibility) are discussed. Cramér's classical theorem on the sum of two independent random variables (for which an 'elementary' proof is much to be desired) is mentioned but not proved or used in this treatment.

In the following chapter the previous theory is generalized to processes whose increments are only orthogonal instead of independent, and this is combined with a detailed discussion of stochastic integrals. An interesting application mentioned is that to Campbell's theorem. However, the application of this theorem to electrical noise is not developed and Campbell's name is not to be found in the bibliography. An interpretation of the idea of a Fourier transform of an actual realization of a process is followed by a generalization and further discussion of stochastic integrals.

Next, in chapter X, we come down to stationary processes with a discrete parameter which are prefaced by a detailed discussion of measure preserving transforms. The strong law of large numbers and the Wold-Khintchine theorem are proved and illustrated and this is followed by a discussion of the effect of linear operators on the spectrum of a process and of processes with rational spectra. In the following chapters all these ideas are discussed for continuous processes.

The final chapter is a rigorous discussion of linear least squares prediction for stationary processes, a subject which is of great interest in practice. The practical applications are, however, not discussed. It is natural to confine the discussion to linear least squares prediction but it is worth pointing out that there is a need for more research on cases where non-linear prediction is better. Consider, for example, a discrete process $\{x_n\}$ where each x_n is distributed uniformly on $(-1, 1)$ and $x_{n+1}=1-2|x_n|$. Then all the serial correlations are zero but an exact predictor always exists. There is also a need for research into processes which are generated by nonlinear difference equations and processes which are not symmetric about their means.

The author is much to be congratulated on this very important book. His aim of giving a rigorous foundation to the subject is, no doubt, mainly responsible for there being no space for a discussion of the problem of inferring the structure of a process from a sample realization, and very little discussion of particular processes. As a result much work on processes is never referred to, and the names of Yule, Bartlett, Pitt, Quenouille and the Kendalls never appear, a fact for which the author rightly apologizes in his preface. The printing, binding, and textual accuracy are of the very highest quality.

Introduction to the Theory of Stochastic Processes Depending on a Continuous Parameter. *Henry B. Mann.* National Bureau of Standards, Applied Mathematics Series, 24. Washington: U. S. Government Printing Office, 1953. Pp. v, 45. $0.30.

ULF GRENANDER, *University of Stockholm*

THIS booklet should be useful for a reader who wishes to find out quickly and not in too great detail what sort of questions are dealt with in the theory of stochastic processes. Its 45 pages contain a good deal of information on this subject.

After a discussion of some basic concepts, the author defines a stochastic process as a one parameter family of stochastic variables and studies various linear operations such as differentiation and integration. Some special processes are discussed in Chapters 2 and 4, mainly of independent increments, and related statistical problems are dealt with in Chapter 3. In Chapter 5 counter data are considered as forming a stochastic process and Chapter 6 finally is devoted to harmonic analysis of processes and the mean ergodic theorem.

The clarity of the exposition and the simplicity of the mathematical machinery that is used makes the book easy to read. The reader who wants to pursue the topic further can find a more complete treatment in two recently published books, J. L. Doob's *Stochastic Processes* (Wiley 1953), and A. Blanc-Lapierre and R. Fortet's *Théorie des fonctions aléatoires* (Masson et Cie, 1953).

Small Particle Statistics. *Gustav Herdan.* Amsterdam: Elsevier Publishing Company, 1953. Pp. xxiii, 520. $12.00.

BENJAMIN EPSTEIN, *Wayne University*

IN MODERN technology, particles ranging in size roughly from 10 down to 10^{-5} centimeters play a very important role. For example, the strength of glassware, ceramic ware, or cement depends to large degree on the fineness and size distribution of the raw material being used. The resistance of dyes and paints to weathering and many other physical properties are strongly affected by the size distribution of the raw materials used and by the way in which they are dispersed throughout the dye or paint. The health of workers in a factory is affected by the kind, density, and distribution of pollutants (generally fine particles) in the atmosphere. In nature, microscopic soil properties are of great importance in sedimentary petrography, in agriculture, and in soil physics. The suitability of coke as a blast furnace fuel can be predicted to some extent from the kinds of size distributions obtained when a sample of coke is broken up into small pieces by the application of various breakage processes. It is virtually impossible to study such properties as "grindability," "resistance to impact," "resistance to abrasion," and the like, without dealing with particle size distributions and how they change, for example, with time or energy expanded.

The author gives an exhaustive account of the current state of knowledge in this field. Since the particles being measured are generally quite small this raises many problems. For example, the problem of how to prepare a sample for measurement, how to carry out the measurements, and what to measure are quite involved technically and important staatistically because of the way in which they can affect the data with which the statistician will be asked to work. The author treats this aspect of the subject admirably, going into such things as sieving, sedimentation, microscopic, and adsorption methods, etc. He also considers various ways of recording data whether by size, by surface area, by weight, etc., and indicates the physical reasons why one measure might be preferable to another depending on circumstances. The reviewer can say, from his own experience, that the statistician called upon to give advice in this field would do well to be aware of such technical considerations.

Roughly half of the book is devoted to a treatment of elementary statistics and the elements of experimental design. This was done by the author in order to make the book self-contained. In the opinion of the reviewer, it would have been better to eliminate most of this material and advise the reader to become acquainted with a basic applied statistics book such as that recently written by A. Hald. Specifically, a good deal of the material in Chapters 2, 4, 6, 7, 8, 10 could have been omitted or treated with greater brevity. By and large the author's treatment of statistics is sound. The author does, however, seem to slip up in the examples on p. 160 and p. 162, since one should surely separate out the effect of variation among the laboratories in running the tests of significance. Analysis of variance is called for and not a simple t-test.

The statistician will find parts of this book very interesting. Among the specially noteworthy features are: (1) critical discussion of how to draw a sample, what and how to measure, kinds of errors likely to appear, etc.; (2) discussion of the distribution laws arising in particle statistics (Chapter 6); (3) discussion of the mechanism of crushing and grinding and associated statistical questions (Chapter 13); (4) statistics applied to problems of mixing (Chapter 14); (5) consideration of the statistics of polymerized materials (Chapter 15); (6) sampling procedure in sedimentary petrology (pp. 417–429). Another good feature of the book are the many illustrations, excellent figures, and extensive bibliography.

Dr. Herdan wrote Chapters 1–17. Chapters 18–23 were written by Dr. M. L. Smith. These latter chapters are devoted to a critical discussion of various experimental methods for determining both size distribution and surface area. This is a complicated problem in the subsieve range and requires very delicate experimental procedures.

To sum up, the book is a "must" for all who work in the statistics of fine particles. It should give the statistician a good deal of insight into the problems peculiar to this field. It should give the technologists and scientists an

appreciation of what statistical methods can accomplish in this area. The book should certainly stimulate healthy cooperation.

U. S. Army, Ordnance Corps: *Tables of the Cumulative Binomial Probabilities.* Ordnance Corps Pamphlet ORDP 20-1, September 1952. Pp. viii+577. 9×12 inches. For sale by Office of Technical Services, Department of Commerce, Washington 25, D. C. at $6.00 per copy. Orders should cite Order No. PB 111389.

THIS is by far the most extensive table of the binomial distribution published yet. Cumulative probabilities are available to seven decimals for population proportions from 0 to 1 by steps of 0.01, for sample sizes through 150 by steps of 1. An Introduction gives a good explanation of the use of the tables, and explains their relations to the Incomplete Beta Function Ratio.

These tables will be indispensable to all practicing statisticians concerned with the binomial distribution.

W.A.W.

The Theory of Inventory Management. *Thomson M. Whitin.* Princeton: Princeton University Press, 1953. Pp. viii, 245. $4.50.

ROBERT DORFMAN, *University of California, Berkeley*

Inventories and their management go back to Joseph, advisor to Pharaoh, at least. The theory of inventory management has a much shorter history, however. Up to the 1920's, inventory policy seems to have been based largely on rule-of-thumb. Some of the simpler problems were formulated and solved during the 1920's, but the development of a systematic theory as a branch of economic and managerial science was undertaken only after World War II, as an aspect of the operations research movement.

Whitin's monograph is an introduction to this promising and fast-growing field. It is divided into three parts. In Part I, Whitin develops the principles of inventory management in the individual firm and compares his conclusions with those of earlier writers, particularly Eiteman and Boulding. Part II deals with the effects of inventory policies on fluctuations in economic activity and the implications of inventory theory for theories of general economic equilibrium. In this connection the theories of Keynes, Metzler, and Leontief are examined in the light of the results of Part I and of empirical data. Part III treats the inventory problems of the national military establishment and applies some of the principles of Part I to them.

The characteristics of an optimal inventory policy for an individual firm, dealt with in Part I, are the foundation of the entire treatment. There are, essentially, two issues to be considered in formulating an inventory policy. The first of these concerns cost minimization: the costs of carrying large inventories have to be balanced against the costs of reordering supplies at frequent intervals. If the cost functions are simple enough, the size of order which minimizes the sum of the ordering cost and the carrying cost per unit

(this is the economic purchase quantity) can be determined by the differential calculus. Whitin shows that this leads immediately to two interesting consequences: first, the economic purchase quantity and therefore the average size of inventories varies in proportion to the square root of the volume of sales; second, the superficial dictum that "the higher the turnover rate the better" is misleading because it may lead to excessively high reordering costs.

The second main issue concerns risk minimization: the risks of overstacking associated with large inventories have to be balanced against the risks of depletion associated with small inventories and against the disorganization of production and sacrifice of sales and goodwill which depletion entails. This is a far more complicated problem than cost minimization. The major aspects of this problem are: (1) the measurement of the losses which would result from overstocking if it occurs and from depletion if that occurs, i.e., the determination of the "loss function" in Abraham Wald's terminology; (2) the estimation of the probability distribution of withdrawals from inventory, which determines the probability of the various possible losses associated with a given inventory policy; (3) the establishment of a policy criterion, be it minimum-maximum loss, minimum expected loss, or whatever; (4) the consolidation of these three aspects into a decision procedure which determines optimal inventory policy as a function of observable data.

Whitin's handling of the risk problem is different in content and spirit from the quadri-partite analysis sketched above. There is little or no discussion of the loss function or of the problems of ascertaining it, and the worked-out illustrations depend on the assumption that the losses resulting from depletion are a known dollar-and-cents unit cost multiplied by the amount of the inventory deficiency. It is assumed without argument that the objective of inventory policy is to minimize the expected loss, so that the issue of selecting a policy criterion does not arise. Nor does Whitin bring up the problem of an integrated decision procedure. Instead he handles separately the two sub-problems of (a) determining the probability distribution of withdrawals from inventory and (b) determining the cost-minimizing policy assuming that the probability distribution is given.

Indeed, Whitin simplifies the problem even further because most of his discussion concerns the case, relatively infrequent in practice, in which the probability distribution is known in advance. In his most extended treatment of a case with an unknown probability distribution, the case of style-goods, he recommends simply that the distribution be estimated by asking buyers to forecast the maximum amount of sales they foresee and adjusting these forecasts in the light of the past performance of the forecasters. (See pp. 70–71.)

What is left after all these simplifications is the problem of minimizing inventory costs and losses, given the probability distribution of withdrawals from stock. The problem is further limited in much of the development by

assuming that the acceptable level of risk of running out of stock has been predetermined. The cost of all these restrictions is suggested by the work of Dvoretzky, Kiefer, and Wolfowitz,[1] who have shown that if reordering costs are appreciable, a policy of a type excluded by Whitin may be optimal.

Whitin's simplifications have the virtue of rendering the inventory problem amenable to the methods of the differential calculus, or the marginal analysis beloved of economists. Even though Whitin's methods would be inadequate in many practical situations, in circumstances in which uncertainty is unimportant and in which the costs of overstocking and understocking can be calculated without difficulty they should provide useful guidance. Moreover, some of the consequences of the analysis, especially the tendency of optimal safety margins to increase according to the square root of the level of sales, are generally valid and highly suggestive.

As applied to the study of economic fluctuations and general equilibrium in Part II, the main consequence of the theoretical analysis is that inventories tend to increase in proportion to the square root of the level of economic activity. Metzler, Boulding, and Leontief[2] have all assumed, at one time or another, that inventories vary in direct proportion to the level of activity. Their theories are, therefore, subject to criticism. Whitin also produces some empirical data which tend to support his position. He concludes, rather tentatively, that although the square-root law mitigates the destabilizing effect of inventories on business cycles, inventories probably do contribute to cyclic instability.

Part III discusses the inventory problems of the national military establishment. The wastefulness of some current rule-of-thumb practices is pointed out in convincing detail. In this context Whitin raises one of the fundamental questions which he neglected in his general treatment: the estimation of the cost of running out of stock of some item or, in other words, the determination of the marginal value of an inventoried item. His proposal is to apply the methods of game theory, that is, to calculate the value of a war game with a given stock of the item and compare this with the value computed with the stock increased by one unit. He nowhere mentions the use of the closely related methods of linear programming which, also, yield estimates of the value of inventories in military and private organizations.

This is the first full-length treatment of a new field. The exposition is generally clear and the mathematics employed are simple and familiar. Many of the results included have already been applied successfully and the treatment unmasks a number of common fallacies about inventory management and behavior. This book can therefore serve usefully as an introduction to the field. But the reader should be warned that this monograph contains

---

[1] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "The Inventory Problem," *Econometrica*, 20 (1952), 187–222, 450–66.

[2] Lloyd A. Metzler, "The Nature and Stability of Inventory Cycles," *Review of Economic Statistics*, August 1941; Kenneth E. Boulding, *A Reconstruction of Economics* (New York, 1950), Part I; W. W. Leontief and others, *Studies in the Structure of the American Economy* (New York, 1953), Chapter 3.

only a smattering of what is known today about the theory of inventory management and that what is known today is only a smattering of what we shall need to know before the theory is ready for wide application.

The Role of Mergers in the Growth of Large Firms. *J. Fred Weston.* Berkeley: University of California Press, 1953. Pp. xvi, 159. $3.50.

See review article by G. Warren Nutter on page 448.

Studies in Income and Wealth, Volume Fifteen. *Conference on Research in Income and Wealth.* New York: National Bureau of Economic Research, 1953. Pp. x, 230. $3.50.

H. S. Houthakker, *Stanford University*

This volume of the well-known series contains eight papers presented in 1950 at a conference in Allerton Park, Illinois. Although the meeting was intended to deal with the distribution of income by size, only one of the papers is strictly concerned with that subject, most of the others being devoted to problems arising in the cross-section analysis of the use of personal income. It must be hoped that a future conference will go into the original, relatively unexplored area, but the quality of some of the papers collected here compensates for the change in emphasis.

The only paper on income size distributions, by George Garvy, is mainly expository. Of greater interest is a contribution by D. Gale Johnson, who tries to elucidate the low incomes of southern farm families by comparing the income of non-farm families in the south and elsewhere. Though hampered by apparent inconsistencies in the data, he advances some remarkable conclusions which are supplemented by comments from other students of regional incomes, who present new data.

In a paper written in 1935, but not hitherto published, Milton Friedman suggests a new method of ranking families of different composition by their relative economic status. Following Sydenstricker and King's pioneering article in the 1920–21 volume of this *Journal*, he proposes the estimation of weights for each class of family members such that, when both income and a particular item of expenditure are divided by the sum of the relevant weights, the resulting relation is independent of family composition. As is recognized by the author and confirmed by calculations in Jean Mann Due's comment this method leads to inconsistent results because the weights depend on the expenditure item considered. It is therefore surprising that Friedman rejects, on "pragmatic" grounds, a more acceptable definition, which distinguishes between income weights and specific weights. The latter approach occurs in R. G. D. Allen's contribution to the Schultz memorial volume (*Studies in Mathematical Economics and Econometrics*, Chicago, 1942) and has been used with some success by authors associated with the

Cambridge Department of Applied Economics (particularly by S. J. Prais in the *Economic Journal* of December 1953).

It is perhaps even more surprising that Professor Friedman should have been so ready to infer the "economic status" of different families from their incomes and expenditures only. Since these data by themselves show only shifts in demand functions, not effects on satisfaction, any such endeavor necessarily involves additional, usually unstated, assumptions, which reflect nothing but the preconceptions of the investigator. The objections to interpersonal comparisons of utility apply with full force here, especially because the presence or absence of children may itself have a strong influence on well-being and this influence cannot normally be isolated.

Much the same point has to be made in connection with Mollie Orshansky's attempt to find "equivalent" levels of living for farm and city dwellers. Except by asking highly sophisticated questions there is no way of determining the income at which city families are as well off as farm families with a given income. The particular method followed in this paper, first suggested by Dorothy S. Brady, is based on the alleged existence of income levels at which the income elasticity of quantity (as distinct from expenditure) reaches a maximum. Quite apart from the statistical difficulties in estimating them, the mere fact that these income levels, if they exist at all, are not the same for all commodities proves that they have no welfare significance whatever. Nevertheless Miss Orshansky's calculations bring out some interesting features of household expenditure patterns.

Data collected by the Survey Research Center of the University of Michigan are analyzed by Janet A. Fisher, who is concerned with the incomes, savings, and assets of families during their economic life cycle. Most of the figures are classified by the age of the household's head, and the remarkable regularities that appear agree fairly well with *a priori* expectations. Further research should reveal whether the use of other indicators of the economic age of the family, such as the duration of the marriage and the number of children, will yield still more informative results, but Miss Fisher's data did not allow such analyses. The importance of this field of investigation is underlined by recent developments in the study of the consumption function, which emphasise its dynamic aspects.

Dynamic factors, though of a different nature, are also considered in an ingenious but obscure contribution by Dorothy S. Brady. By comparing budget surveys from various periods she tries to identify a "normal form" of the consumption function, in which not only the level of but also the change in income is regarded as a determining variable. Not having any data on income changes at her disposal, she apparently allows for the latter effect by using community income as well as family income. The precise logic of her method is unfortunately not made clear. It might be suspected, for instance, that in comparing the 1917–19 with the 1935–36 survey changes in the price level should be taken into account, but nothing is said on that topic.

Nor is the reader's bewilderment allayed by cryptic footnotes such as "Expenditures and savings were standardized to an average family size of 3.5 persons." It is quite possible that Mrs. Brady's approach leads to important results, but the present exposition is mystifying rather than convincing.

Margaret G. Reid's valuable paper illustrates J. R. Hicks' dictum to the effect that the income one can calculate is not the income one seeks, whereas this true income escapes calculation. She starts from the observation that the income elasticity of consumption is less for farm families than for others, and analyzes to what extent this may be due to the income concept used rather than to differences in behavior patterns. The choice of concept is particularly important in the case of farm families because of the difficulty of separating operating and living expenses and the related problems of depreciation and inventory changes. Similar questions arise for all families with highly variable incomes. Miss Reid surveys some alternative classifications of families with reference to the resulting biases of this nature, but she defers a definite recommendation.

In a concluding paper Simon Kuznets outlines directions of further inquiry. His stimulating suggestions, which demonstrate a penetrating insight both in what is most necessary and in what is feasible in empirical research are too numerous to be discussed here. He might have put a little more emphasis on theoretical research, however; even if this does not lead to immediately applicable conclusions it may yet help to clarify the issues and methods involved. Several of the contributions to this volume would have been improved if the models used had been more explicitly formulated. As it stands the volume shows impressively what a wealth of basic information is already in existence; the main problem for the near future is how to exploit it more effectively. Despite their various, and mostly excusable, defects these papers constitute a significant advance in the solution of that problem.

Better Population Forecasting for Areas and Communities. *Van Beuren Stanbery.* (Domestic Commerce Series No. 32). U. S. Department of Commerce. Washington, D. C.: U. S. Government Printing Office, 1952. Pp. iv, 80. 25 cents. Paper.

FREDERICK F. STEPHAN, *Princeton University*

THE problem of making forecasts of population growth for cities, counties, metropolitan districts frequently arises in the work of local industries, public utilities, zoning and planning agencies, and municipal organizations concerned with the provision of schools, water, sanitation and other services. Frequently they find it necessary to plan far ahead because their facilities must be built in single units that can not be enlarged without excessive expense. Forecasts of local population growth are affected greatly, of course, by the factors that influence the location of new industry and the expansion of business and other factors that affect employment in, and migration of

population to or from the area. In addition to these factors, there, are of course, the natural growth of the resident population and the progression of each cohort through the ages at which it enters the labor force, establishes new households, and otherwise affects the total demand for the services for which the population estimates are being prepared.

If the lot of the forecaster of population growth for the entire nation is a difficult and unhappy one—and recent experience would seem to make this undeniable—then the life of a forecaster of local populations must certainly be unbearable. For such a hazardous occupation it would seem that any accumulation or accretion of knowledge and any tricks or devices no matter how crude, would be helpful. While it is too much to hope that a highly scientific methodology could be put together for forecasting populations in all kinds of communities, there are some communities in which special factors do permit highly accurate forecasting if full advantage is made of them. Nonetheless, many statisticians look with misgiving at the relatively crude methods of estimation that must be employed, perforce, by anyone so brash as to attempt to make local population forecasts.

The booklet that is under review is a contribution to "those who use or make population projections." It reviews some of the more familiar methods of forecasting, such as the projection of past population growth, projection by use of the relation to national or regional areas, for which forecasts are already available, projection of migration and natural increase, and projection from specific estimates of future employment. Work sheets are sketched for these computations. Much advice is given about things to consider. It is largely a manual of suggestions and cautions written in unsophisticated language.

This guide book will be welcomed by many who want to see "better population forecasting." Perhaps it is not without its liabilities, however, for it may encourage mediocre work at the expense of really good analysis. The bibliography lists a number of studies that would have increased the value of the text if they had been used adequately as examples. Moreover, there is little evidence that the author has looked into the forecasting that is done by the telephone and utility engineers or some of the regional business research bureaus. The methods of projection could well be improved by use of data correlated with population growth, such as meter installations, to fill in the gaps between censuses and to permit more effective linkage to economic projections for the area. If this is done then more efficient statistical methods can be used. All this suggests that local estimates should be prepared with the advice and assistance of those who can contribute all the data and techniques and judgment that are available in the community and pertinent. Such a group would be likely to emphasize the width of the band of uncertainty that cloaks the forecasts, an aspect that this booklet very properly does too.

Population Changes in Europe Since 1939. *Gregory (Grzegorz) Frumkin.* New York: Augustus M. Kelley, Inc., 1951. Pp. 191.

DOUGLAS F. DOWD, *Cornell University*

THIS study would appear, at first glance, to be one of modest compass, dealing with assuredly useful but nevertheless dreary data. In Mr. Frumkin's hands the data come alive, his task is revealed as a staggering one, and one leaves the book shocked anew by at least one meaning of Hitler's War: the death and displacement of scores of millions of people.

Mr. Frumkin—now with the United Nations, and editor of the *Statistical Year Book of the League of Nations* throughout its existence—attempts "a systematic determination of the magnitude of the changes which have occurred in Europe's population since 1939 and of their determining factors . . . country by country, on a uniform pattern, by means of balance sheets . . . " (p. 9).

In the brief first chapter, the population background of Europe (before 1939) is discussed, separate charts for twenty countries showing birth and death rates and the natural increase of population are presented, and we are furnished with a handful of pithy remarks and caveats hinging on the demographer's craft, but worth pondering by all those who study society: e.g., "there is sometimes more stability behind changing figures and greater mobility behind stable figures than one would imagine" (p. 19).

Chapter II provides a lucid and interesting discussion of the balance sheet approach of computing population changes. This approach—"somewhat arduous, strictly inductive"—the reviewer found easy to follow, clear, and highly informative. The method may be illustrated by listing the items in the French balance sheet, 1939–1945: Population, end of 1938; Births; "Normal" deaths; War losses (military; civilians, non-Jewish; Jews killed); Population shifts, net balance; Population end of 1945. Similarly, with some different inclusions, for 1946–1947. Items are tallied plus or minus.

Chapter III is the heart of the book. Here we find detailed balance sheets for twenty-four European countries, and a careful discussion of relevant data for the U.S.S.R. The distinction between the latter and the former rests on the non-availability of statistics from the U.S.S.R. Each item on the balance sheet is systematically explained, the reliability of the figures discussed, and statistical conclusions are drawn. One cannot fail to be impressed by the meticulousness of Mr. Frumkin's research, and his success in achieving an "unbiased enquiry". On a different note, one cannot read these pages without becoming sickened by the violence and destruction and misery underlying the figures.

Chapter IV summarizes the results of the study, in a valuable combined balance sheet for all countries, and goes on to say something of the meaning of the figures and the prospects for the future. Some of these remarks are worth quoting here: "In spite of heavy losses, Europe . . . emerged from the turmoil of war with a larger population in 1947 than on the eve of the war."

(p. 175). "A characteristic feature of the last war was . . . that the main loss of population was due not to fighting, but to mass-murder by the German invader . . . Civilian losses were overwhelmingly concentrated in areas occupied by the Germans, and the number of Jews murdered made up almost one-half of the total civilian losses. In Poland alone the number of Jewish victims exceeded 3 millions" (p. 182). "War and genocide . . . accounted . . . for the death of 15 million persons, almost 6 millions being military deaths, and over 9 million deaths among civilians" (p. 181). Add to this the 17 millions of war losses estimated for the U.S.S.R., which Frumkin estimates as being "on the low side" (p. 164). Add the millions in the "exceptionally large shifts of the population chased across national boundaries" (p. 188). The book concludes with a grim warning: "As the result of World War II, national minorities in Europe have mostly disappeared. Mass-murders, shifts of frontiers and ruthless mass-transfers were the instruments by means of which the former European mosaic has been converted into an array of ethnically homogeneous units each with the sign: 'Trespassers will be prosecuted.' These units, born under coercion, cannot be maintained ethnically pure except through coercion" (p. 190).

Mr. Frumkin throughout emphasizes the tentative nature of his conclusions. Tentative though they may be, the diligence, the years of experience, and the considerable talents of Mr. Frumkin which come through on every page would seem to indicate that it will be Frumkin or those following his methods who will improve upon what looks to be a definitive work.

Bristol on the Move—A Travel Survey. *British Transport Commission.* London: 1953. Pp. 46. 10s.6d.

Leonard P. Adams, *Cornell University*

This study was authorized by the British Transport Commission and carried out by Research Services Ltd. in the winter of 1950–1951. It was designed to provide information on the travel patterns of the people of Bristol (defined for present purposes to include those residing within a ten-mile radius [more or less] of the city) of interest to the sociologist, to the operator of public services, and to the advertiser. The principal findings with respect to the methods used in travel, costs to the traveler, types of travelers (age, sex, income class, etc.), purposes of travel, time required, and other items are presented in a series of tables supplemented by a narrative discussion and some excellent photographs of the city and its environs.

No doubt the data obtained from this survey have been of service to transport operators and advertisers, and probably to some extent to sociologists. From the standpoint of those not particularly interested in the Bristol area per se but interested in local studies of travel patterns, the design of the Bristol survey and the methods used will be of interest. This reviewer has recently been studying commuting patterns of industrial workers, so is probably inclined to be more critical of some of the methods used than those

interested in administration or selling, although in some significant respects, as will be noted, the study has weaknesses from their point of view also.

In designing the study Research Services Ltd. evidently assumed that there was little value in explaining the geographic limits of the Bristol area. This assumption may be well founded in fact but there is no explanation of why the ten mile limit was chosen or why the possibilities of longer distance travel to or from Bristol were ruled out. The authors admit that there is considerable traffic between Bristol and Bath which has been excluded from the study because the City of Bath "has transport services of its own which would have to form the subject of a separate investigation." But there is no way of telling from the report whether or not some of the households surveyed have closer economic ties to Bath than to Bristol even though they live within the ten mile radius from Bristol. Data on travel are not given in terms of places of origin and destination. Omission of such information, together with a predetermined definition of the Bristol area, raises questions about the scope of the area and the relationship of its people to other population centers that the published summary does not answer. To what extent do the people of Bristol have economic and social ties with those in Bath? How much cross traffic is there? For purposes of advertising and labor supply marketing, should Bath and Bristol be combined in a single area?

Presumably the investigators considered that for their purposes a more or less arbitrary definition of the Bristol area would be adequate. In any case, it facilitated the selection of the sample households to be visited. The primary sample was chosen by taking every 150th address from the Electoral Registers covering the sample area. A secondary sample consisting of the fifth address following each primary address was also selected at the time the primary was drawn. When the householder at the primary address could not be reached or would not cooperate the nearby secondary sample was substituted. In case there were two or more households at the same address, additional interviews up to three were taken in order to avoid weakening the geographic distribution of the sample. Such additional interviews were called "subsidiary" interviews. Each member of the household was questioned concerning his or her use of public transport in the seven days preceding the interview. The methodology used was, in general, the same as that followed in the London Survey, except that *all* use of public transportation in the preceding seven days was recorded instead of just the "regular" journeys. Although no measures of the adequacy and reliability of the sample are given, it seems probable that the results give a reasonably good picture of the characteristics of travel within the area selected.

While a single cross-section view of travel patterns has value from the standpoint of operators and advertisers, it has severe limitations if one is trying to understand *trends* in housing location, the journey-to-work, mode of travel and other matters. Travel patterns in the Bristol area, if they are at all similar to those in industrial centers in this country, are probably dy-

namic, changing with such factors as new housing developments and the location of new industrial plants—and, judging by some of the photographs in the report, the Bristol area has had some interesting post-war housing developments. Changes in travel patterns probably could be measured with a fair degree of accuracy by the interview method. Experience with travel studies in this country suggests, however, that employer personnel records provide information more readily with respect to journey-to-work patterns and they also show changes over time. If similar records are available in English plants, Research Services Ltd. in making future studies may wish to use this source to provide a basis for defining the geographic limits of the area to be surveyed and also to show changes in the distances traveled and characteristics of the work force employed.

International Shipping Cartels. *Daniel Marx, Jr.* Princeton, N. J.: Princeton University Press, 1953. Pp. xiii, 323. $6.00.

ROCKWOOD CHIN, *Berea College*

HERE is a scholarly book about shipping conference agreements as a form of international cartel. This topic is usually neglected by books on international cartels and inadequately treated in transportation texts. The author is to be commended for his comprehensive description of the nature, history, purposes, and methods of shipping cartels. Competition from non-conference liners and tramp shipping is also discussed.

Particularly detailed is his analysis (in Chapters IV, V, and VII) of American and British experience in the investigation and regulation of shipping practices dating from about the beginning of the twentieth century. All the main modern arguments for and against shipping cartels are virtually covered in the early reports of the Royal Commission on Shipping Rings (1909) and the Alexander Committee (1914). Out of the recommendations of the latter there evolved in the United States the Shipping Act of 1916 which, among other things, prohibited deferred rebates, "fighting ships," and unjust discrimination. Curiously enough, there is no specific reference to the Bland-Copeland Bill of 1935 which unsuccessfully attempted to reverse the 1916 legislation.

Chapter VI includes brief uneven sketches of restraints imposed upon national participation in shipping conferences by other selected countries—British Dominions and some European and Far Eastern countries. The economics of shipping conferences, discussed in Chapters II, III, and XII, supplemented by a note in the Appendix on discriminatory pricing, provide a theoretical background for the factual presentation.

The reader may jump to Chapter XIV for the conclusions of the book without loss of the main argument. In trying to be objective, the author sometimes appears to vacillate in his attitude toward international shipping cartels, and there is occasional ambiguity as to what he is definitely recommending. Generally speaking, he regards international shipping cartels as

inevitable, capable of abusing their privileges, and also of suffering from wastes of competition. Monopoly and cutthroat competition are both decried. "The record seems to indicate that relatively few conferences have been guilty of very serious abuses, unless exclusive patronage contracts and pooling arrangements approved by the Commission are so condemned" (p. 136). Regulation by national governments and an international agency (at the time of writing, the Intergovernmental Maritime Consultative Organization had not yet come into existence) would both have their limitations, though desirable for supervising tying arrangements and for preventing monopoly abuses and undue discrimination. Some sort of improvement in self-regulation through the existing conference system is favored, but the author does not make clear how this can actually be attained.

Chapter III, on the general economic and political environment, deals with the relation of shipping to international trade, location of economic activity, and balance of payments. It contains the bulk of the book's trade and shipping statistics. Elsewhere statistics are few. Though there is a table in Chapter XII comparing fluctuations in indices of tramp freight and conference general cargo rates, actually freight rates are cited rather infrequently. In Chapter IX there are also charts showing the names and membership of various freight and passenger conferences participating in United States foreign trade and travel. One misses, however, the complete text of a conference agreement of the type described in the book, of which we are told there are over one hundred on file at the Maritime Commission (now the Federal Maritime Board).

A word of caution on figures given as shipping company earnings of foreign exchange: The early British figures (p. 43) really refer to total earnings by British shipping companies in carrying exports *and imports*, as well as inter-third country trade, and are therefore not all additions to foreign exchange. Of course, the British method of c.i.f. valuation of imports compensated for the overstatement of transportation credits in the balance of payments. The 1949 shipping earnings figure, calculated on an f.o.b. basis, is not comparable to the earlier figures. Similar remarks apply to the Norwegian data cited on p. 42. Furthermore, in singling Norway out as the country for whose economy shipping is of the greatest importance, the ratio of shipping foreign exchange earnings to national income would have been a more significant criterion than net earnings figures alone or their ratio to the trade balance.

On the whole this is a very fine book on a much-neglected subject. A wealth of information has been pieced together from books, documentary reports, government agencies, and case studies. It fills a need in the literature of international cartels and transportation. Students of economics, government, and transportation will find it very useful. Since the author can already point out some comparisons with the international aviation field, I hope that he may someday also publish a treatise on international air transport cartels on the same high level.

PUBLICATIONS RECEIVED

Actuaries, Society of. *1951 Impairment Study.* 1954. $7.50.

Arizona, University of, Agricultural Experiment Station. *Barriers to the Interstate Movement of Milk and Dairy Products in the Eleven Western States.* Tucson, April 1954. Paper.

Barclay, George W. *Colonial Development and Population in Taiwan.* Princeton: University Press, 1954. $5.00.

Blackwell, David, and Girshik, M. A. *Theory of Games and Statistical Decisions.* New York: John Wiley and Sons, 1954. $7.50.

Bovet, Eric. *L'organisation rationelle de la distribution moyen de stabilisation économique.* Paris: Delachaux et Niestlé, 1954. Frs. 12.50. Paper.

Burington, Richard Stevens, and May, Donald Curtis, Jr. *Handbook of Probability and Statistics with Tables.* Sandusky, Ohio: Handbook Publishers, Inc., 1953. $4.50.

California, Department of Industrial Relations. *Union Labor in California, 1953.* San Francisco, 1953. Paper.

Cameron, Burgess. *The Determination of Production.* New York: Cambridge University Press, 1954. $3.75.

Cavé, R. *Le control statistique de fabrications.* Paris: Eyrolles, 1953.

Chevalley, Claude C. *The Algebraic Theory of Spinors.* New York: Columbia University Press, 1954. $3.75.

Clem, Mary A., and Federer, W. T. *Random Arrangements for Lattice Designs.* Ames: Iowa State College Press, 1950. $1.00. Spiralbound, paper.

Federer, W. T. *Random Arrangements for Some Three-Dimensional Lattice Designs.* Ithaca, N. Y.: Cornell Agricultural Experiment Station, 1953. $2.50. Spiralbound, paper.

Furmin, Peter A. *The Michigan Business Receipts Tax: An Appraisal.* Ann Arbor: University of Michigan, 1953. $2.00. Paper.

Ghosh, M. K., and Prekash, Om. *Principles and Problems of Industrial Organization.* Allahabad: The Indian Press, Ltd., 1953. Rs. 11.

Gilbert, Milton, and Kravis, Irving B. *An International Comparison of National Products and the Purchasing Power of Currencies: A Study of the United States, the United Kingdom, France, Germany, and Italy.* Paris: OEEC, 1954. $3.00.

Henderson, W. D. *Britain and Industrial Europe, 1750–1870: Studies in British Influence on the Industrial Revolution in Western Europe.* Liverpool: University Press, 1954. 25 s.

Herskovits, Melville J. *Franz Boas: the Science of Man in the Making.* New York: Charles Scribner's Sons, 1953. $2.50.

Hoel, Paul G. *Introduction to Mathematical Statistics. Second Edition.* New York: John Wiley and Sons, 1954. $5.00.

Hooker, P. F., and Longley-Cook, L. H. *Life and Other Contingencies, Vol. 1.* New York: Cambridge University Press, 1953. $4.50.

International Labour Office. *Yearbook of Labour Statistics, 1953. Thirteenth issue.* Geneva, 1953. $5.00. Paper.

Liebenstein, Harvey. *A Theory of Economic-Demographic Development.* Princeton: University Press, 1954. $4.00.

Lieblein, Julius. *A New Method of Analyzing Extreme-Value Data.* Technical Note 3053. Washington, D. C.: National Advisory Committee for Aeronautics, 1954. Paper.

Macdonald, Gordon D. *Apartment Building Construction Manhattan 1902–1953.* New York: The Real Estate Board of New York, Inc., 1953. $5.00. (To members of the National Association of Real Estate Boards and other professional associations $1.00.)

Macdonald, Gordon D. *Office Building Construction Manhattan 1901–1953, and supplement number one.* New York: The Real Estate Board of New York, Inc., 1953. $10.00. (To members of the National Association of Real Estate Boards and other professional associations $5.00.)

Marcus, Edward. *Canada and the International Business Cycle, 1927–1939.* New York: Bookman Associates, 1954. $3.75.

Marvick, Dwaine. *Career Perspectives in a Bureaucratic Setting.* Michigan Governmental Studies No. 27. Ann Arbor: University of Michigan, 1954. $2.25. Paper.

Mauldin, W. Parker, and Akers, Donald S. *The Population of Poland.* Washington, D. C.: U. S. Government Printing Office, 1954. $1.00. Paper.

Maxwell, E. A. *An Analytical Calculus, Vol. I, Vol. II.* New York: Cambridge University Press, 1954. $2.75, $3.50.

Mitchell, Robert B., and Rapkin, Chester. *Urban Traffic: A Function of Land Use.* New York: Columbia University Press, 1954. $5.00.

Modley, Rudolf, and Cawley, Thomas, Jr., editors. *Aviation Facts and Figures.* Washington, D. C.: Lincoln Press, Inc., 1953.

National Bureau of Standards. *Tables of*

Circular and Hyperbolic Sines and Cosines for Radian Arguments. Applied Mathematics Series 36. Washington, D. C.: U. S. Government Printing Office, 1953. $3.00.

———. *Tables of Lagrangian Coefficients for Sexagesimal Interpolation.* Applied Mathematics Series 35. Washington, D. C.: U. S. Government Printing Office, 1954. $2.00.

National Industrial Conference Board, Inc. *The Economic Almanac, 1953–1954.* New York: Thomas Y. Crowell Company, 1953. $2.95.

Neifeld, M. R. *Trends in Consumer Finance.* Easton, Pa.: Mack Publishing Company, 1954. $6.00.

New York, State of, Department of Labor. *Employment Patterns of Insured Workers in Selected New York Industries, 1947–1951.* New York: Division of Research and Statistics (1440 Broadway), 1953. Paper.

Phelps, Clyde Williams. *Instalment Sales Financing: Its Services to the Dealer.* Baltimore, Md.: Commercial Credit Company (14 Light St.), 1953. Paper.

Prest, A. R., assisted by A. A. Adams. *Consumers' Expenditure in the United Kingdom, 1900–1919.* New York: Cambridge University press, 1954. $7.50.

Richardson, C. H. *An Introduction to the Calculus of Finite Difference.* New York: D. Van Nostrand Co. Inc., 1954. $2.50.

Salzer, Herbert E. *Tables of Coefficient for the Numerical Calculation of Laplace Transforms.* Applied Mathematics Series 30. Washington, D. C.: U. S. Government Printing Office, 1953. 25 cents. Paper.

Schutzenberger, M. P. *Contribution aux applications statistiques de la théorie de l'information.* Vol. III, Fascicules 1–2, 1954. Paris, Institut Henri Poincaré. Paper.

Schultz, William J., and Reinhardt, Hedwig. *Credit and Collection Management. Second Edition.* New York: Prentice-Hall, Inc., 1954. $9.00.

Spurr, William A., Kellogg, Lester S., and Smith, John H. *Business and Economic Statistics.* Homewood, Ill.: Richard D. Irwin, Inc., 1954. College price, $6.00.

Tanganyika. *Report on the Census of the Non-Native Population Taken on the Night of 25 February 1948.* Dar Es Salaam: Government Printer, 1953. 25/. Paper.

Tax Institute, Inc. *Federal-State-Local Tax Correlation, a symposium volume.* Princeton, N. J., 1954. $5.00.

Uganda Protectorate. *Report on the Census of the Non-Native Population of Uganda Protectorate Taken on the Night of 25th February, 1948.* Nairobi, 1953. Paper.

United Nations, Department of Economic Affairs. *Economic Survey of Asia and the Far East, 1953.* Bangkok, February 1954. $1.50. Paper.

———. *Economic Survey of Europe in 1953 including a Study of Economic Development in Southern Europe.* Geneva, 1954. $2.50. Paper.

———. *A Study of Trade between Asia and Europe.* Geneva, 1953. $1.50. Paper.

United Nations, Department of Social Affairs, Population Division. *Additional Information on the Population of Tanganyika.* New York: 1953. 30 cents. Paper.

United Nations, Economic Commission for Asia and the Far East. *Mobilization of Domestic Capital Report and Documents of the First Working Party of Experts.* New York: Cambridge University Press, 1952. $1.50. Paper.

———. *Mobilization of Domestic Capital: Report and Documents of the Second Working Party of Experts.* Bangkok, 1953. $2.50. Paper.

United Nations, Economic Commission for Europe. *Annual Bulletin of Transport Statistics 1952.* Geneva, 1952. $1.25. Paper.

United Nations, Statistical Office of. *Concepts and Definitions of Capital Formation.* Statistical Papers, Series F, No. 3. New York: 1953. 25 cents. Paper.

———. *Demographic Yearbook 1953.* New York, 1953. $5.00, paper; $6.50, cloth.

———. *International Migration Statistics.* Statistical Papers, Series M, No. 20. New York, 1953. 25 cents. Paper.

———. *Population and Vital Statistics Reports.* Statistical Papers, Series A, Vol. VI, No. 1. New York, 1954. 30 cents. Paper.

———. *Retail Price Comparisons for International Salary Determination.* Statistical Papers, Series M, No. 14 add. 1. New York, 1953. 30 cents. Paper.

———. *Statistical Yearbook 1953.* New York, 1953. $6.00, paper; $7.50, cloth.

———. *Statistics of National Income and Expenditure.* Statistical Papers, Series H, Nos. 4 and 5. New York, 1953. 60 cents and 80 cents. Paper.

———. *A System of National Accounts and Supporting Tables.* Studies in Methods, No. 2. New York, 1953. 50 cents. Paper.

United Nations, Technical Assistance Programme. *The Economic and Social Development of Libya.* New York, 1953. $1.75. Paper.

Vicary, James M. *An Annotated Bibliography of Word Association References to Marketing Researchers.* New York: James M. Vicary Company (22 E. 60th St.), 1954.

Free. Mimeo.

Waugh, Frederick V. *Reading on Agricultural Marketing.* Ames: Iowa State College Press, 1954. $5.00.

Wert, James E., Meidt, Charles O., and Ahmann, J. Stanley. *Statistical Methods in Educational and Psychological Research.* New York: Appleton-Century-Crofts, Inc. 1954. $5.00.

Whelpton, Pascal K. *Cohort Fertility: Native White Women in the United States.* Princeton: University Press, 1954. $6.00. Offset.

Wilks, S. S. *Analisis Estadistico Elemental.* Traduccion de Sixto Rios y Jose Royo. Madrid: Consejo Superior de Investigaciones Científicas, Departmento de Estadistica, 1952. Paper.

Wold, Herman. *A Study on the Analysis of Stationary Time Series. Second Edition.* Stockholm: Almquist and Wiksell, 1954. Sw. kr. 38.

Wolfenden, Hugh H. *Population Statistics and their Compilation.* Chicago. University Press, 1954. $7.50.

Woytinsky, W. S., and Woytinsky, E. S. *World Population and Production: Trends and Outlook.* New York: The Twentieth Century Fund, 1953. $12.00.

Žarković, S. S., *Greške popisa stanovništva.* (*Population Census Errors.*) Beograd, 1954. Paper.

RANDOM DIGITS (17,376–20,875)

From *A Million Random Digits* to be published for The Rand Corporation toward the end of 1954 by The Free Press, Glencoe, Illinois.

04808	99531	47991	46064	80467
71924	64882	94893	82935	99076
56410	89552	28404	74525	74212
38851	16144	99542	27481	21992
91428	10589	09454	43308	66753
40083	17141	30702	31997	69856
93419	10474	41796	88285	02448
03704	65516	65448	20203	21189
78181	90060	74904	42627	16638
45972	93572	76011	03426	50226
60898	63968	62264	64603	51866
40398	54180	65869	87977	02799
68245	76912	01222	59516	36438
27019	15248	66444	25267	05171
99868	88894	43769	52239	05919
87904	74135	53842	59520	23979
68851	41049	97190	53984	04773
71742	57223	66599	86071	01901
02742	48803	17823	22093	43907
56181	96052	67211	61712	54590
55355	61548	55988	47309	23749
78961	41072	09876	18903	30292
92654	97226	53434	77025	63892
13757	37719	84450	02697	60309
05776	85945	74651	00216	50842
71039	83083	60427	78495	99809
61672	01184	46438	27698	40652
42988	77983	58708	42176	67356
13652	16640	27896	26907	86760
53186	97859	97213	19859	41037
47890	10690	26486	38744	25943
65654	34629	88831	97253	67282
00324	17120	39900	67135	42772
48244	26191	88421	90491	83290
64081	47704	15018	45600	17241
60617	06414	56596	63011	24193
72860	18452	42983	23931	11789
04631	55283	19605	34163	86540
06884	15444	09310	17048	24243
26611	09551	82626	38104	58432

58932	89326	33491	04617	88092
73073	52171	89301	74066	82717
42665	80748	77622	15779	37361
59985	59807	60562	85747	94028
50943	40422	63035	60344	06883
22224	02627	91576	16781	89184
24473	42096	76920	88864	54164
38582	21871	14672	93362	67981
46094	43845	91838	79574	08003
91061	31674	73729	99315	16699
00397	56753	53158	71872	68153
14328	44708	72952	27048	67887
88534	87112	68614	83073	88794
97347	87316	73087	77135	71883
01366	72976	01868	51667	63279
37106	20523	21584	93712	83654
06476	70603	97122	44978	78028
81717	48410	94516	15427	85323
51583	69788	41758	55004	30992
50120	33884	83655	88345	69602
89761	23053	77480	28683	68324
08943	66660	11057	98849	29499
71685	97247	79368	43710	80365
17402	66300	94385	01717	96191
52606	39860	92127	42588	93307
66035	07223	76264	29148	68652
21565	30786	45403	33782	93424
88735	75275	03080	77653	55430
50404	80166	28017	52611	60012
80834	11317	93109	91857	47904
26872	72927	79021	51571	68825
16530	96086	17329	87959	23727
84644	00448	86828	50552	84832
88620	72894	94716	84622	49771
22209	78590	68615	58113	23727
04795	53971	14592	39634	63855
54291	56045	61635	32186	86651
30654	48543	18339	65024	33386
11123	08732	49393	12911	75803
56577	51257	83291	12329	17827
58987	02026	42969	59144	84349
16851	99197	70476	77113	46320
02104	49435	77706	18924	24957
54440	07893	31618	35707	65130
87681	42543	69847	81848	32034
24337	61634	52574	83649	28725
62557	25292	72781	17186	10393
02913	03885	58822	82941	43806
68706	87619	13846	56197	27151
05930	33213	78416	00194	91369

70119	31347	12659	11574	70052
98390	30240	28330	41145	16918
08172	23823	48433	57222	34435
21238	19051	50768	40807	88681
79342	44640	93942	97371	16842
93039	79367	00812	41365	04515
62865	09576	97207	33739	78345
00800	72496	24767	61768	07228
64340	02224	48336	14891	72188
92168	52692	31224	12185	43065
20494	18813	16242	40257	66402
87693	30242	70545	69128	51528
05567	05561	82071	07234	67690
85166	37189	75671	33879	27411
26704	47922	56650	40236	66207
01047	81624	77395	62310	41501
58183	21952	84098	28913	55736
64667	57092	21315	04731	71877
27149	13843	09817	09407	88276
66232	80293	74502	36925	60184
40500	21406	00571	87320	81683
35892	49668	83991	72088	30210
54819	26094	51409	21485	94764
64224	47909	09994	23750	17351
36913	58173	45709	83679	82617
64254	64745	10614	86371	43244
82018	25536	74031	31807	70133
28833	44043	96215	21270	59427
96879	27659	95463	53847	40921
95938	76014	99818	16606	19713
97154	71237	06073	57343	51428
78790	17026	59008	28543	11576
25034	59325	08844	95774	49323
70116	44091	88505	15575	44927
66904	23000	73259	68626	98962
91171	28299	62619	81550	46798
74547	13260	79262	55831	83784
30448	14154	75795	39465	82353
06584	29867	45898	66415	89349
68548	86576	14344	75889	04514
49319	50206	22024	56124	50749
81034	86779	34622	70859	33045
68905	44234	18244	31602	38388
88530	72096	44459	31449	93182
37227	11302	04667	32526	64713
83220	50529	20619	11606	10297
66703	30017	35347	35038	16648
69556	76728	60535	59961	76979
99040	96390	65989	38375	30332
85185	72849	58611	31220	66108

// JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 268 DECEMBER 1954 *Volume 49*

METHODS OF APPORTIONING SEATS IN THE HOUSE OF REPRESENTATIVES

Walter F. Willcox*
Cornell University

The subject of Congressional Apportionment was transferred a few years ago to the Judiciary Committees of Congress; no member of either committee, I believe, was in Washington in the late twenties when the long fight over the question ended with the dangerous increase in the size of the House halted and with reapportionment made a ministerial act about which Congress need no longer concern itself after each census as it had done for 130 years.

In view of that situation the Judiciary Committee of the House asked me to report to them on methods of apportionment and to include so much of the congressional history of the subject as might help them in dealing with it. My reply to the committee was sent some months ago; it ended by proposing two amendments to the automatic act of 1929, one of them changing the present method to a novel one which I now think the best, the other lopping off ten members automatically after each future census from an overgrown House until it approaches a membership of three hundred, the number mentioned often in congressional debates as a desirable goal. This article will supplement my report to Congress and lay before scholars the main results of my work which has lasted for more than half a century, a period during which some of my early conclusions have changed and several new ones emerged.

I had to prepare apportionment tables for Congress after the census of 1900. After the following census two groups of students entered the field. One, for whom Allyn A. Young spoke, held that the question of method "is mathematical and ought to be decided upon the basis of a general consensus of mathematical opinion."[1] The position of the other group was stated thus: "The question involved in the apportionment of

* Editor's Note: Professor Willcox was President of the American Statistical Association in 1912 and of the American Economic Association in 1915.

[1] Allyn Young, circular letter to statisticians and mathematicians dated Nov. 19, 1926 (unprinted).

representatives is primarily one of constitutional law. The role of mathematics in the problem is to make clear the consequences of any given interpretation of the Constitution."[2]

This article will defend the second position. My understanding of the difference between the two groups was stated in a letter to an opponent: "Our conference left me with the impression that all the men present except Professor X and me believed that there is only one right method of apportionment, the method of equal proportions. . . . My view is that the methods can be arranged in an order of preferability, that their sequence in that order depends upon the predominant object to be secured by apportionment, that the object of apportionment is a political rather than a mathematical problem and one to be determined, therefore, not by academic students but by Congress."[3]

Those who wrote the Constitution intended, I believe, to make the resident of a state the unit of representation in the House, as they had made the state itself the unit in the Senate. The Constitution contains three passages which bear on the method by which to carry out that intention. They are:

(1) "The number of Representatives shall not exceed one for every thirty thousand."

(2) "Representatives shall be apportioned among the several States according to their respective numbers."

(3) "Each State shall have at least one Representative."

The first passage alone determined the method of apportionment used by Congress before 1840, the second alone has underlain the four methods used since 1840, the second modified by the third furnishes a basis for the novel method proposed herein. My arguments are addressed primarily to Congress because in deciding this question Congress is the jury, but I cannot succeed in that quarter without support from the public, hence this article.

The House of Representatives at first had only 65 members, but Congress soon became convinced that it should be enlarged as much as possible. The limit on size set in the first passage proved to be ambiguous; for five months Senate and House disputed over how to interpret it. Did it mean not more than one for every thirty thousand in each state, 112 in all, or not more than one for every thirty thousand in all the states, 120 in all? The House sent on to the Senate a bill apportioning 112 representatives. The Senate returned it after adding eight seats

[2] F. W. Owens, "On the Apportionment of Representatives," *American Statistical Association Publications*, 17 (1921), p. 968.

[3] Walter F. Willcox, circular letter of Aug. 12, 1931 (unprinted).

for large remainders and in that form it reached the President. Washington vetoed it, relying on the advice of Jefferson, then Secretary of State and in charge of the census, that the method was unconstitutional. That veto established the method which Congress used for half a century; it may be called the Jefferson method or method of rejected fractions.[4] Its essential feature is that it apportions representatives only for units.

This method was abandoned forty years later as a result of Webster's criticism. Under the discarded method the larger a state the smaller the proportion between its rejected fraction and its population, hence the smaller its district population and the more representatives it would get. Under 1950 figures, for example, the method of rejected fractions would have transferred to states above the average size nine seats which the present method gave to states below the average size.

Among the tests of a method one needs to be mentioned here because it underlies Webster's method. First divide the states into three groups, large, small and very small, the line separating large and small being the average population of those two groups, and the line separating small and very small being whether a state gets its one seat for population or by constitutional guarantee. Next compute the average district population of each of the two groups of large and small states. The nearer those averages are to each other, the better the method.

This test also measures the nearness to equal representation given to one person or one million persons wherever the residence; that is probably what those who wrote the Constitution had in mind.

Webster's proposal was to give supplementary seats for fractions larger than one half, "major fractions." An apportionment bill had passed the House after the 1830 census and been referred to his committee. He reported that the method of rejected fractions was unconstitutional because it did not meet the second constitutional requirement and apportion seats to the several States according to their respective numbers. Webster computed the number of representatives a few states would have if the proportion each had of 240 was as near as possible to the proportion of its population to that of all the states. He showed that the population of New York and Vermont, for example, would entitle the former to 38.59 representatives and the latter to 5.65 and claimed that New York should receive 39 seats and Vermont 6 instead of the 40 and 5 given them in the bill.

If he had been able to show Congress that the current method must

---

[4] Jefferson wrote in a memorandum for which Washington had asked, "Fractions must be neglected because the Constitution . . . has left them unprovided for." Wrtings. (1940 ed.), vol. 3, pp. 201–11.

overrepresent a large state and underrepresent a small one, his argument might have carried greater weight. He might have said that by it the average district population in large states would be much below that in small ones, and that, if Congress should adopt his amendment, the difference would be only one sixth as great. Stated in a way perhaps more meaningful to Congress, his method would have transferred seats from three large states, Kentucky, Pennsylvania, and New York, to three small ones, Delaware, Missouri and Vermont.

He carried the Senate, but when the House rejected his amendment the upper house yielded. His failure may have been due even more to a weakness in his mathematics which I have explained elsewhere and which I was able eighty years later[5] to correct.

Webster's contributions were that he called attention to the second constitutional requirement, showed that the results of the current method violated it and that to apportion supplementary seats for fractional remainders larger than one half would be a great improvement. But he did not answer either of two questions the first of which long vexed Congress. That is, How is the common divisor which his method needed to be found? The other is, How is the problem affected by the third requirement in the constitution?

In the period between 1832 and 1910 Congress tried to adapt Webster's revolutionary idea to a problem the shape of which was changing. In 1842 Congress experimented with his method. The law specified the common divisor which he did not know how to compute for a specified number of seats and provided for "one additional Representative for each State having a fraction greater than one moiety of the said ratio,"[6] but it did not specify the size of the House. When the 1850 census was at hand Congress was minded to stabilize the size of the House at 233 members by a ministerial apportionment. The law instructed the Secretary of the Interior, first, to divide the combined population of the states by 233, then to divide the population of each state by the quotient, and finally to apportion one seat for each unit and enough supplementary seats for large fractions to reach the required total.

This method dodged one of the difficulties in Webster's proposal, how to find the common divisor, but two others remained. The method might yield one or more seats for large minor fractions, or withhold seats for one or more small major fractions. As long as the House did not increase in size the method worked well, but after the 1870 census

[5] Walter F. Willcox, "Last Words on the Apportionment Problem," *Law and Contemporary Problems*, vol. 17, p. 291.

[6] *Statutes at Large*, vol. 9, p. 432.

Congress resumed the policy of apportioning enough seats to keep every delegation intact and the experience of forty years with that change wrecked the method. The Superintendent of the Census began to send Congress tables based on the 1850 method which showed the distribution of each number of seats within the limits of size which interested the House. These tables showed that now and then one or more states might receive a seat for a large minor fraction, one or more might fail to receive a seat for a small major fraction, or when one member was added to the total two states might gain and a third lose a seat (the "Alabama paradox").

As I have said, my connection with the problem began in 1900 when I was placed in charge of a division in the Census Bureau which was to prepare the apportionment tables. The law said we should use the Vinton method, but, as tables based on earlier figures had shown its weakness, we submitted a second set based on an imperfect understanding of the Webster method.

The sizes of the House which interested Congress were 357, its existing size, and 386, the size at which no state would lose a seat. Our Vinton tables contained two examples of the Alabama paradox at just these numbers, one affecting Colorado, the other Maine. For Colorado, the figures ran:

Size of House	*356*	*357*	*358*
Seats for Colorado	3	2	3

For Maine they were:

Size of House	*384*	*385*	*386*	*387*
Seats for Maine	4	4	3	4

Congress got over the hurdle and apportioned 386 seats by starting with the table for 384, which contained two quotients with major fractions for which no seats were apportioned. For each of these quotients Congress gave an extra seat and thus reached 386, the number desired, with no state losing a seat and no major fraction unrewarded.

After following the long House debate I returned to Cornell sure that the principle of the Webster method was sound but its mathematics weak. Once the difficulty was grasped, the solution was obvious; apply the sliding divisor concept. After the 1910 figures had been announced I took to Washington a set of tables and an explanatory letter in which I had written:

"The history of reports, debates, and votes upon apportionment seems to show a settled conviction in Congress that every major fraction gives a valid claim to an additional Representative. . . . The pres-

ent method is based upon that conviction and seeks to facilitate action in conformity with it. Because of this feature I have called it the method of major fractions.

"The results are simple, but the method itself is somewhat difficult to explain. If a ratio of 240,000 persons to each Representative be assumed arbitrarily as a starting point, that number divided into the population of each state and one Representative assigned for each whole number and each major fraction in the series of quotients, a total of 383 Representatives is reached. If the ratio be then diminished by 10 to 239,990, no difference in the apportionment will result, but the decimal in each quotient will be slightly increased. If the ratio be further reduced to 239,980, 239,970, etc., the decimals continue to increase with each change of ratio, but with varying rapidity. It is a simple problem to compute in which State the decimal will first pass .500 and become a major fraction and at just what ratio the change will occur. In the present case the State whose decimal first reached .500 is Illinois and the corresponding ratio is 239,940, . . . which has been called the boundary ratio.[7]

The Bureau of the Census handed Congress two other sets of tables, one based on the prescribed Vinton method, the other on a novel mathematical analysis of the problem devised by my successor in Washington and perfected later by Professor E. V. Huntington. The 1911 apportionment was based on the Cornell tables.

The figures of the 1920 census showed that the Huntington method, or method of equal proportions, would give three seats to small states, (Rhode Island, Vermont and New Mexico) which the Webster method would give to large ones (New York, North Carolina and Virginia). This difference in the results was due to the fact that the Huntington method makes the "critical fraction" separating large remainders for which seats are apportioned from small ones for which they are not a variable one lying always below .500 and above .414. The following computation shows why I preferred the results of the Webster method.

States	*Deviation from Population Standard by*	
	Webster method	*Huntington method*
3 large states	+1.15	−1.85
3 small states	−1.41	+1.59
Total	2.56	3.44

The total deviation from the standard is one third greater by the Huntington method than by the Webster method.

[7] 61st Congress, 3d Session, House Report No. 1911, Jan. 13, 1911, p. 9 f.

The two methods applied to 1930 figures yielded identical results. Ten years later Congress abandoned the Webster method and adopted the Huntington one largely because it would give to Arkansas a seat which the Webster method would have given to Michigan. After the 1950 census the Huntington method gave to Kansas a seat which the Webster method would have given to California. Analysis of these instances shows how the two tests worked.

Professor Huntington described his test in these words: "*Test of equal proportions.*—A transfer of a seat from one state to another should be made if, and only if, the percentage difference between the congressional districts[8] in the two states would be reduced by the transfer."[9]

This leads to the following figures (in thousands)

State	*Before transfer* No. of seats	*Before transfer* District Population	*After transfer* No. of seats	*After transfer* District Population
California	31	353	30	341
Kansas	5	381	6	318
Disparity	8.0 per cent		7.5 per cent	

Because 7.5 per cent was less than 8.0 per cent, the Huntington test gave the seat to Kansas. I cannot see how that test bears on the problem it tries to solve.

If we compute the number of seats and fractions of a seat to which the two states were entitled in 1950 California should have had 30.68 and Kasas 5.52. The Huntington method in giving California 30 seats curtailed its representation by .68 and in giving Kansas 6 seats swelled it by .48, a total departure from the standard of 1.16. The Webster method in giving California 31 seats would have swollen its representation by .32 and in giving Kansas 5 would have curtailed it by .52, a total departure from the standard of .84, about three fourths as much.

Another form of comparing the methods may be clearer. Each method gave the two states together 36 representatives. The question is, should California have received 31 and Kansas 5, or California 30 and Kansas 6? Since California had 84.75 per cent of the population of the two states it should have that per cent of the 36 seats, in other words 30.51 seats for it and 5.49 for Kansas; so California had the stronger claim to the transferable seat.

Cumulative evidence comes from applying each test to the 1940 fig-

[8] He meant by congressional districts what I prefer to call district populations.

[9] E. V. Huntington, "Methods of Apportionment in Congress," in 76th Congress 3d Session Senate Document No. 304, p. 3.

ures for Michigan and Arkansas. Each method gave those states in combination 24 seats; the question was, should Michigan receive 17 and Arkansas 7, or Michigan 18 and Arkansas 6? Michigan was entitled by its population to 17.43 seats and Arkansas to 6.04 so Michigan with 17 was curtailed by .43 and Arkansas with 7 was strengthened by .96, a total deviation of 1.39. But if Michigan had received 18 seats and Arkansas 6 the total deviation would have been only .61, (.57 and .04) less than half as much. Michigan had 72.95 and Arkansas 27.05 per cent of the population of the two states, so the former was entitled to 17.51 and the latter to 6.49 out of the 24 seats. Evidently Michigan should have been given the transferable seat.

The argument thus far has indicated that Congress erred when it adopted the Huntington method and abandoned that of Webster.

We come to the last question, What effect has the requirement, "Each State shall have at least one Representative" on the problem? That question but not its answer I saw through a glass darkly when I wrote, "The Vinton method . . . involves a fundamental theoretical error.[10] It overlooks the crucial fact that seats in the House of Representatives are of two classes, the 48, one for each state, which are guaranteed by the Constitution and are as completely beyond the control of Congress as the seats of the Senators are, and the remainder, the number and distribution of which are under congressional control. The two classes might be named the apportionable and the unapportionable seats. The fact that they are not individually distinguishable has apparently been responsible for the failure to recognize their existence. To get this theoretical requirement clearly in mind it may be helpful to think of the seats in the House of Representatives as numbered. The first 48 seats, one for each state, would be numbered one to indicate that there is no basis for distinguishing between them. The next seat, numbered 49, would be apportioned to New York, number 50 to Pennsylvania and so on".[11]

This distinction has now been recognized by all authorities and apportionment tables give the states to which seats go in succession from No. 49 on to No. 435. Both the present and the proposed method under 1950 figures would give seats 49, 50, 51, and 52 to the largest states, New York, California, Pennsylvania and Illinois, in the order of their size but differ about seat 53. The present method gave that seat to New York with a district population (in thousands) of 7,415 although

---

[10] Later I realised that the Webster method in either form and the Huntington method involve the same error.

[11] Walter F. Willcox, "The Apportionment of Representatives; Annual Address of the President," *American Economic Review*, 6 (1916) Supplement, pp. 3–16.

Ohio, the fifth state in size, had a larger district population, 7,947, and in my judgment should have received the seat. If Congress should agree with me on that point, its choice would entail a decision that the method of included fractions leads to better results than the method of equal proportions and should displace it at the first opportunity.

In ending the argument about methods of apportionment we may summarize the conclusions.

Methods are of two kinds, primary and secondary, the former finding a root in the Constitution, the latter not finding such a root.

There are three primary methods, one based on "not more than one in every thirty thousand," another based on "according to their respective numbers," and a third based on the attempt to combine the second requirement with the last, "Each State shall receive at least one Representative."

The results of these three methods differ in that the first apportions no supplementary seats for remainders, the second apportions such seats for remainders larger than one half, the third like the first apportions no seats for remainders but unlike it assigns one seat to each state before apportionment begins.

The results also differ in their distribution of transferable seats, the number of which has increased from two after the first census to sixteen after the last. The first method distributes all transferable seats among the large states, the second divides them as evenly as possible between the large and the small states, the third distributes them all among the small states.

The Constitution seems to leave with Congress a choice between two tests and two methods of apportionment. The first test would be based on the nearness of two proportions, on the one hand the proportion that a state's population makes of the population of all the states, and on the other the proportion that the number of a state's representatives makes of the whole number of representatives. The second test would be based on the nearness of the district populations of the 48 states to one another.

The number of methods of apportionment giving different results varies with the figures of a census but is always one more than the number of transferable seats. The seventeen results possible after the 1950 census would come from three primary and fourteen secondary methods; these seventeen methods make up a series wherein the results of each would differ from those on either side of it by transferring one seat from a large state to a small one or vice versa.

The methods reaching these results use a series of seventeen divisors

with 329,577 at one end and 365,394 at the other. They use also a series of seventeen critical fractions with zero at one end and one at the other.[12]

The article may close with a few words about another amendment now before the House Judiciary Committee. It provides for a slight automatic reduction in the size of the House after each future census. The law now declares that the President shall report to Congress after each census the result of redistributing the then existing number of members, among the states according to the method last used by Congress. The amendment would insert the words, "ten less than" before the words, "the then existing number."

The House is now about four and a half times as large as the Senate; at the start it was only two and one half times as large. In state legislatures the difference between the size of the two branches is much less, and, since as a rule the more recent a state constitution is, the nearer to equality in size are the upper and lower houses, it would seem that American experience has led to a reduction of the average difference.

More important evidence comes from members of the House. Sixteen years ago a Congressman acting at my suggestion asked each of his colleagues whether he was satisfied with the present size of the House and, if not, whether he wanted it larger or smaller. Among the one third who replied, one half were satisfied with the present size of the House; of the other half about nine tenths wanted it smaller.

More evidence to the same effect came from the apportionment debate in the 1920's; then fifteen experienced and influential representatives gave an opinion; all but two wanted the House smaller. Probably no one now a member can recall, as Congressman Burton then could, when it was only three quarters as large. He had begun his long service more than forty years before and had been for six years a Senator. He said: "I began when there were 325 members of this body, and the disadvantages in the transaction of business now as compared with then are beyond my powers to describe. It is not only the greater expense but the greater confusion on the floor and the greater difficulty in the orderly transaction of work. . . . I would rather see this House consist of 300 members than 435."

With the size of the House stabilized as it has been for fifty years, 1910–1960, the decennial increase in the average population of a con-

[12] "The Apportionment Problem and the Size of the House; A return to Webster," *Cornell Law Quarterly*, vol. 35 (1950), pp. 367–89; and "Last Words on the Apportionment Problem," *Law and Contemporary Problems*, vol. 17 (1952), pp. 290–301.

gressional district is about 44,000. If the size of the House had been reduced by ten members after the last census, that increase would have been about 49,000, a difference probably no member would worry about.

If such an amendment should be adopted the business of the House might be done faster and better and debating the amendment, even if Congress took no action on it, would bring home to it the fact that it can now change the size of the House slowly up or down without being blocked as it often was in the past by a tiny pressure group.

THE KINSEY REPORT ON FEMALES*

Dorothy S. Brady
Washington, D.C.

The first chapter of "Sexual Behavior in the Human Female" contains four pages of persuasion in three sections—"The Scientific Objective," "The Right to Investigate" and "The Individual's Right to Know." The argument begins in the first section with three cautiously phrased sentences—

> It should be clearly understood that the original goal of our study was the extension of our knowledge in an area in which scientific information appeared to be limited. In the course of the years it has become apparent that the data we have acquired may prove of value in the consideration of some of our social problems, but that was not why we originally began this research.
>
> It has been the history of science that any addition to our store of adequately established knowledge may ultimately contribute to man's mastery of the material universe. (p. 7)

progresses to the emotional level in the second section—

> The scientist who observes and describes the reality is attacked as an enemy of the faith, and his acceptance of human limitations in modifying that reality is condemned as scientific materialism. But we believe that an increased understanding of the biologic and psychologic and social factors which account for each type of sexual activity may contribute to an ultimate adjustment between man's sexual nature and the needs of the total social organization. (p. 10)

and ends in the third section with a dramatic defense of scientific freedom and individual rights—

> . . . We believe that the scientist who obtains his right to investigate from the citizens at large, is under obligation to make his findings available to all who can utilize his data. Any scientist who fails to report or to place his findings in channels where they may serve the maximum number of persons, fails to recognize the sources of his right to investigate and thereby jeopardizes the rights of all scientists to investigate in any field. . . .
>
> . . . We believe that if we have any right to investigate in this field, we are under obligation to make the results of our investigations available to all who can read and understand and utilize our data. (pp. 10, 11)

The obligation to make the findings available to the maximum number of persons was fulfilled with all the craft and skill of modern publicity. Yet the message conveyed to the public is not at all unequivocal.

* A review article on *Sexual Behavior in the Human Female*, by Alfred C. Kinsey, Wardell B. Pomeroy, Clyde E. Martin, and Paul H. Gebhard. Philadelphia and London: W. B. Saunders Company, 1953. Pp. 842; $8.00.

As a sociologist put it, "Even the careful reader, trying to avoid selective bias, is not always sure where these investigators stand. They frequently hint at adverse evaluations of our religious-moral traditions but draw back in the end, leaving evaluative issues for future study. This somewhat confused situation is providing a field day for moralists. Both outraged traditionalists, scanning these books for purple passages, and opponents of religion, eager to prove the validity of their preconceptions, are finding what they want to find."[1]

The issues surrounding the individual's right to know the results of publicly sponsored investigations should be considered apart from the subject of the research. The social control of scientific research ultimately relies on professional codes of practice. Neither naïveté nor indifference but plain common sense on the part of the general public and their representatives in legal office or on boards of foundations delegates the responsibility for research to the scientists as a group. There are not many important subjects of research that can be presented so that all men, women and children could pretend to "read, understand and utilize the data." Medicine is replete with examples of research on subjects of such grave importance to the general public that progress towards scientific findings makes the headlines of daily newspapers. Yet the sifting of contradictory evidence, the synthesis of generalizations, and the validity of applications are trustingly left to the medical profession. Kinsey's interpretation of the code of the scientist,

> . . . to investigate honestly, to observe and to record without prejudice, to observe as adequately as human sense organs or the most modern instruments may allow, to observe persistently and sufficiently in order that there may be an ultimate understanding of the basic nature of the matter which is involved. (p. 10)

stops short of the principle which safeguards the public interest in scientific research.

Modern science and its technological applications grew out of the concept of an experiment that can be repeated by different observers. Historians tell us that this idea in its time was revolutionary enough to create great public interest. The assurance that the results of an experiment will be repeated again and again is fundamental in the application of the results of scientific research. The results of an experiment are quantitative generalizations. The mere recording of observations has seldom led to a conflict between the observers and the legal or

[1] Claude C. Bowman, "Sociological Implications of the Kinsey Studies," Temple University, paper presented to the Eastern Sociological Society, April 4, 1954.

religious authority. The conflicts have centered around the quantitative generalizations that constituted attacks on whole systems of ideas. The scientists with whom Kinsey compares himself—Kepler, Copernicus, Galileo, and Pascal—were not primarily observers; they generalized astronomical observations that had been accumulating for thousands of years.

Astronomy in the nineteenth century met a single historical event, the problem of the observation, that was difficult, if not impossible to replicate. Towards the end of the century as the mass of scientific data and generalizations became large and complex and empirical methods were extended to the biological and social sciences, the concept of replication had to be given a new operational definition equivalent to its literal meaning.

Modern statistical procedures give the answer to a serious question. What can be done to assure the validity of the results of observations that, literally, cannot be repeated? It is no historical accident that Karl Pearson, one of the founders of modern statistics, tried to reformulate the nineteenth century concept of the repeated experiment in his *Grammar of Science*. It is no historical accident that his son, Egon Pearson, among others, brought statistical procedures to the testing of hypotheses. Statistical theory offers means for appraising the results of investigations that serve the function that the actual repetitions of the experiment generally served in the past and still serve in many provinces of scientific research. A profession recognizes the work of one of its members after scrutiny of all the operational phases of the investigation. Statistical methods are involved in the appraisal at the primary level of sampling and at the level of quantitative generalization.

The specification of basic definitions and concepts is not within the province of statistical methods. It is the random or systematic variability in observations of particular phenomena or in reporting on particular forms of behavior that the statistician is equipped to study. The formulation of the operating hypotheses on the basis of which experiments or surveys are designed is also not a function of the statistician. Statistical procedures implement a principle expressed nearly a century ago—"Wrong hypotheses, rightly worked, have produced more useful results than unguided observations."[2]

Kinsey's haste to present his findings to all those who might want to apply them has forced into the popular press a process usually conducted quietly through technical journals and conferences. The public may be diverted by all the controversies over the validity and impor-

[2] DeMorgan, "A Budget of Paradoxes," Open Court Edition, p. 87.

tance of the Kinsey findings but large groups of specialists in the theoretical and applied fields concerned with the subject matter are understandably bewildered.

The use of statistical survey methods in the study of sex behavior did not originate with the Kinsey research. Kinsey's volumes refer to earlier studies and Appendix B of the report of the American Statistical Association's committee[3] summarizes important differences between methods used in the earlier surveys and in the Kinsey survey. The data from a survey in a related field, "Social and Psychological Factors Affecting Fertility,"[4] which was conducted in Indianapolis in 1941, and was also financed by a foundation, have been carefully analyzed in more than twenty technical articles by demographers and sociologists. Many of the original hypotheses formulated for testing with the data have been rejected. The final summaries present the results of all the statistical tests. The data from the Indianapolis survey and from the other surveys of sex behavior do not display the neat and systematic differences or similarities between population groups that Kinsey and his associates have discovered in their study of the human female.

The Kinsey group places great emphasis on its interviewing techniques and approach to the respondents. The evaluation of these techniques in the report of the Committee of the Association and in Wallis' review[5] do not need to be summarized in this connection. All of the questions about the sample of males also relate to the sample of females which is, " . . . even more inadequate than our sample of males in representing lower educational levels, rural groups, and some of the other segments of the population" (p. 57). Yet the relationships shown in the analysis of female sex behavior are uniform and regular compared to the results from most large-scale surveys of much less variable types of human behavior. Other investigations, even with longer and much more intensive statistical analysis, do not produce such elegant statistical results.

Here lies the paradox. What is the Kinsey secret that, from an inadequate sample, produces results of a stability that others cannot duplicate? Even if the original reports by individuals interviewed were absolutely accurate the mathematical uniformities in the correlations

[3] Appendix B of the report of a committee appointed in 1950 by S. S. Wilks as President of the American Statistical Association to review the statistical methods used by Alfred C. Kinsey, Wardell B. Pomeroy, and Clyde E. Martin in "Sexual Behavior in the Human Male" (Philadelphia, W. B. Saunders Co., 1948), to be published in a monograph by the American Statistical Association in 1954.

[4] Clyde V. Kiser and P. K. Whelpton, "Résumé of the Indianapolis Study of Social and Psychological Factors Affecting Fertility," *Population Studies*, Vol. 7, No. 2, November 1953, Great Britain.

[5] W. Allen Wallis, "Statistics of the Kinsey Report," *Journal of the American Statistical Association*, 44 (1949), pp. 463–84.

would be surprising. When there is a significant difference in sex behavior between groups differentiated by some social characteristic there are few cases of inversions in the rankings. When there is no correlation the absence of a significant relation can scarcely be questioned.

Most social surveys offer a multiplicity of hypotheses about the connection between a particular form of human behavior and the demographic, social, and economic characteristics of various population groups. The choice of the hypotheses to be tested, evaluated, and presented to the general public more and more has to be made by the investigating agency with such advice from specialists in the subject matter and in statistical methods as it can marshall on the problem of selection. It is doubtful whether the best hypotheses in a social survey would ever yield as many unambiguous comparisons as the Kinsey study of the human female. Certainly those hypotheses—and they are not necessarily the best—that lead to published data do not reach the Kinsey standard.

The particular detail in the Kinsey tables that does not conform to experience with other empirical data is the virtual absence of zero frequencies in the incidence tables that are published. When frequencies of a particular form of behavior run low—under 15 per cent of the population group—other empirical studies often show a zero frequency in the sample reports for subgroups of the population even when the number reporting in that subgroup is substantial—say between 50 and 500. The statistical reason for a wide variation in the percentages and averages in samples so large is to be sought in the theory of multivariate distributions. Most social and economic characteristics of the population are highly intercorrelated. The accidents of sampling may result in a subgroup defined by one or two factors that differs widely from the total population represented by that subgroup with respect to other factors.

The uniformities of the Kinsey correlations and the acknowledged variability in sexual behavior lead to only one conclusion. Most of the Kinsey findings must be based on a few, relatively incontrovertible relations. The apparent precision of his results must be based on something that is effectively a tautology.

When the factors associated with a particular form of behavior are correlated with each other, a single simple and perhaps logical relation can be reproduced indefinitely through one-factor correlations that appear to be highly significant. The Kinsey report on the human female finds a systematic inverse connection between sexual behavior of females and religious activity and a lack of relation with educational

attainments which were found significant in the study of males. Both of these results may be merely reflections of a correlation of age and marital status with religion and education among the women interviewed by Kinsey and his associates.

The report on the human female provides in the first chapter some information on the basic sample not offered in the volume on the human male. With the information on the distribution of the sample by religious groups and age at the time of interview and the cumulated tables in the rest of the volume, it is possible to trace some of the intercorrelation of the religious classification and age and—less completely—the connection between religious activity and marital status. Of the devout females, Protestant, Catholic, or Jewish, relatively more were under 25 years of age than of the moderately active or inactive. The proportion of devout Protestant females in the age bracket 16–20, was 32 per cent and approximately twice the percentage of inactive Protestants of the same age. Among devout Catholic and Jewish females there were 33 and 52 per cent in that age bracket; among inactive 19 and 39 per cent. The difference in the age distributions among the groups classified by the degree of religious adherence means that more older females were reporting on their activities in the younger ages among those inactive religiously than among the devout. The average age of those reporting on activity at a given age differs among the groups classified by religious background much as their reported sexual activity differs.

The cumulated tables reveal that there were more females married in the younger age groups among those who were inactive than among those who were moderate or devout in their religious activities. With some difficulty it is possible to read another difference among the religious groups from the tables presented in this volume. Relatively more previously married females appeared among the inactive than among the moderate or devout religious groups.

All of the relations between religious affiliation and sexual behavior may be a reflection of more fundamental association between age, marital status, and the frequencies of various forms of sexual experience reflected in the Kinsey samples. The connection between sexual experience and marital status in the case of females can hardly be debated and Kinsey's results can easily be confirmed by the man on the street who has read the Bible and observed the social practices in his own time. The connection with age probably can not be explained as a cultural tautology.

The report on the human female shows changes in the female be-

havior occurring in this century in the nineteen-twenties that have persisted until the present time. The relation between sexual behavior and decade of birth, and by inference to age, brings back that very inconvenient problem of the statistical sample. The Kinsey sample of females was constituted mainly of women who had been to college, at least for a few years, and whose parents had the capacity to finance girls as well as boys through college. The girls who went to college prior to World War I must have differed more from the general population of females than those whose parents were able to support this particular luxury during the twenties. The Kinsey relation between decade of birth and female sexual behavior may be only a reflection of the more representative selection of the total female population that entered colleges and universities after the close of World War I. The popular literature of that decade stressed the appearance of the girl in search of a husband on the college campus. If, prior to the war, the woman who went to college was selected by her own drive towards a professional activity or attainment, her behavior in other respects may have been very atypical.

The Kinsey report on the female comes to a sweeping conclusion about the educated women still single in their older years:

> When such frustrated or sexually unresponsive, unmarried females attempt to direct the behavior of other persons, they may do considerable damage. There were grade school, high school, and college teachers among these unresponsive or unresponding females. Some of them had been directors of organizations for youth, some of them had been directors of institutions for girls or older women, many of them had been active in women's clubs and service organizations, and not a few of them had had a part in establishing public policies. Some of them had been responsible for some of the more extreme sex laws which state legislatures had passed. Not a few of them were active in religious work, directing the sexual education and trying to direct the sexual behavior of other persons. Some of them were medically trained, but as physicians they were still shocked to learn of the sexual activities of even their average patients. If it were realized that something between a third and a half of the unmarried females over twenty years of age have never had a completed sexual experience, parents and particularly the males in the population might debate the wisdom of making such women responsible for the guidance of youth. (p. 526)

In all there were 299 unmarried females 36 years of age or older and of these, 112 were 46 or older.

With so many of them teachers, directors of institutions, and medically trained as the text of this report suggests, it is possible to reach only one conclusion. These were the women of an earlier era dedicated to a professional life and—perhaps of more importance in the present

context—were among those who "covered up" their reports on experience. The struggle of women for professional recognition in many fields is over. The particular type of determined individual who entered professional training before 1918 and may have been represented by the few that came into Kinsey's sample can not be considered representative of women of the same age in the years to come.

Whether the relation between age, decade of birth, and sexual behavior relates only to Kinsey's sample or to a more general group of the female population, differences in age and marital status explain most of the correlations found between sexual behavior and religious activity. In other words, it appears that sexual activity explains religious inactivity. The smoothness of the Kinsey correlations relating sexual activity inversely to religious activity is apparently the regularity imposed by dividing a range of a variable into three parts, low, medium, and high.

Had the authors standardized their comparisons for age and marital status, and had marital status been defined in four groups to include the status of "being engaged," all of the generalizations in this volume about the connection between sexual behavior and social factors might have been changed.

Kinsey and his associates apparently started with a fairly simple hypothesis about the sexual behavior of human females—not too well formalized to be sure—that it has an anatomical and physiological similarity with the males, but is inhibited by social, legal, and other cultural codes and practices.

Through the series of tables describing the sexual activities of the religious groups of females there is a causal chain that deflected Kinsey and his associates from their initial and relatively simple operating hypothesis. The greater sexual activity of the religiously inactive females appears to lead inevitably to a higher proportion of broken marriages. The Kinsey inference that greater sexual experience for the female before marriage leads to quicker and more satisfactory sexual marital relations is contradicted by two correlations in his survey. Pre-marital hetero-sexual experience is correlated with extra-marital experience. The groups that have the highest incidence of extra-marital experience include the greatest number of marriages broken by separation and divorce.

The authors contend with these inferences from their study in various oblique ways. Early in the text they state: "The generalizations throughout the present volume have therefore been restricted to the particular sample that we have had available." Progressively to the

final chapters they throw away this restriction. The book would make an excellent subject for textual analysis by the techniques of historians and Biblical scholars.

> There are some who have feared that a scientific approach to the problems of sex might threaten the existence of the marital institution. There are some who advocate the perpetuation of our ignorance because they fear that science will undermine the mystical concepts that they have substituted for reality. But there appear to be more persons who believe that an extension of our knowledge may contribute to the establishment of better marriages. (p. 13)
>
> There are legal and social responsibilities in any marriage; there are economic problems to be solved; above all, there are psychologic adjustments to be made between the wedded partners. Sexual adjustments represent only one aspect and not necessarily the most important aspect of marriage. No balanced program for American youth can be confined to preparing them for sexual relationships in marriage. But it is inconceivable that anyone who is objectively and scientifically interested in successful marriages should fail to appreciate the significance of coitus in marriage, or wholly ignore the correlations which exist between pre-marital activities and the sexual adjustments which are made in marriage. (p. 391)
>
> These correlations between pre-marital and extra-marital experience may have depended in part upon a selective factor: the females who were inclined to accept coitus before marriage may have been the ones who were more inclined to accept non-marital coitus after marriage. A causal relationship may also have been involved, for it is not impossible that non-marital coital experience before marriage had persuaded those females that non-marital coitus might be acceptable after marriage. (pp. 427–428)
>
> Extra-marital coitus had figured as a factor in the divorces of a fair number of the females and males in our histories. We have data on 907 individuals (female and male) who had had extra-marital experience and whose marriages had been terminated by divorce. We have the subjects' judgments of the significance of their extra-marital coitus in 415 cases. In nearly two-thirds (61 per cent) of these cases, the subject did not believe that his or her own extra-marital activity had been any factor in leading to that divorce. . . . It is to be noted, however, that these were the subjects' own estimates of the significance and, as clinicians well know, it is not unlikely that the extra-marital experience had contributed to the divorces in more ways and to a greater extent than the subjects themselves realized. (p. 435)
>
> These data once again emphasize the fact that the reconciliation of the married individual's desire for coitus with a variety of sexual partners, and the maintenance of a stable marriage, presents a problem which has not been satisfactorily resolved in our culture. It is not likely to be resolved until man moves more completely away from his mammalian ancestry. (p. 436)
>
> The failure to recognize these differences in the needs of the two sexes for a regular sexual outlet may be the source of a considerable amount of difficulty in marriage. It is the source of many social disturbances over questions of sex. In establishing sex laws, in considering the sexual needs of females and males in penal and other institutions, in considering the need among

females and among males for non-marital sources of sexual outlet, and in various other social problems, we cannot reach final solutions unless we comprehend these considerable differences between the sexual needs of the average female and the average male. (p. 682)

The possibility of reconciling the different sexual interests and capacities of females and males, the possibility of working out sexual adjustments in marriage, and the possibility of adjusting social concepts to allow for these differences between females and males, will depend upon our willingness to accept the realities which the available data seem to indicate. (pp. 688–689)

The reality the authors found, perhaps reluctantly, in the last chapter, namely, a fundamental difference between males and females, is supported by information from the survey that shows "the male's greater inclination to be promiscuous" (p. 683). The table shows the number of partners reported by males and females in pre-marital petting and pre-marital coitus. The difference in average numbers of partners in pre-marital coitus reported by females and by males is so great that, in view of the fact that the numbers of males and females in the population are nearly equal, only one conclusion can be reached: that the samples of males and females came from different populations. This possibility is admitted in the first chapter in a discussion of the sample of females. The discrepancy in the data on this subject, summarized differently, is explained on page 79 in fine print. Reasons given, among others, are differences in the distribution of the samples by educational attainment, the omission of prostitutes from this report, the fact that some of the men reported on experience abroad while in the armed forces, and the possibility that "the females may have covered up in reporting their pre-marital experience, or the males may have exaggerated their reports of such experience."

The inferences in the last chapters represent the confusion of investigators not equipped with technical tools, without much experience in the analysis of multivariate relations. At some point in the analysis of their data for females, the authors were forced to reject their original hypothesis. Without a specific formulation of the quantitative relations being tested, they may not have recognized what happened. The importance of social and psychological factors as developed mainly in the final chapters and references in earlier chapters could easily have been interlarded.

The book leaves the very strong impression that even Kinsey and his associates would not replicate the generalizations in this volume. The contradictions in their inferences will serve them well. Several other systems of generalizations can be shown to be consistent with some of the pronouncements in this volume.

UNSOLVED PROBLEMS OF EXPERIMENTAL STATISTICS*

JOHN W. TUKEY
Princeton University

IT WOULD not be misleading to suggest that there is really only one unsolved problem of experimental statistics: "How can we recognize the problems of experimental statistics?" We can recognize a good many unsolved problems by accident, but we probably miss many important ones for far too many years. Difficulties in identifying problems have delayed statistics far more than difficulties in solving problems. This seems likely to be the case in the future, too.

Thus it is appropriate to be as systematic as we can about unsolved problems. Any system may be a start toward, or even a partial solution of, this problem of recognition. I shall try to do this by stating first some principles and then some consequences. I shall strive to phrase all these principles as generally as possible, in the hope of prolonging their useful life.

A discussion of examples of these 18 general principles will set forth a certain number of unsolved problems, while a list of 51 provocative questions poses many more. (This list is admittedly and intentionally incomplete.) The account closes with a discussion of the possibility of orienting experimental statistics toward problems rather than techniques.

SOME GENERAL PRINCIPLES

If we feel that the detailed problems of experimental statistics arise from the interaction of certain general principles among themselves and with classes of experiments, it is reasonable to try to state and illustrate some of these principles. Before stating the hypergeneral principles on which these general principles hang, we need to explain the sense in which three terms, *ends*, *areas* and *considerations* will be used there and in the sequel.

By an *end* we refer to real purposes of the user of the statistical technique. These purposes are often unformulated, and their partial formulation often requires the statistician to "psychoanalyze" his client (in the writer's view this is one of the most important functions of the statistical consultant!). An *immediate end* is a formalized (and almost certainly partial) end such as to describe an appearance (e.g., by a point estimate), to make a test of significance, to make a decision, or to reach a confidence statement.

* Prepared in connection with research sponsored by the Office of Naval Research. Presented to the American Statistical Association and the Biometric Society 28 December 1953.

An *area* is a class of situations with qualitatively similar data, such, for example, as the class where two sets of observations are presented for the comparison of the "typical" values of the corresponding populations (means, medians, and the like serve as "typical" values). Within an area, different techniques are competitive. Within an area, the historical, evolutionary, and logical relations of different techniques are relatively clear.

A *consideration* is a recognition that the world may very well be more complex, annoying, and difficult than our earlier techniques had supposed. Thus we might admit—nay, even take into consideration—the possibility that we did not know the variance, that the distribution might not be normal, that a certain fraction of the observations are affected by blunders, etc.

The four hypergeneral principles, which may seem harmless until we come to their consequences, run as follows:

(*A*) Different ends require different means and different logical structures.

(*B*) In each area, statistical method must and does evolve, mainly by adding *both* immediate ends *and* considerations.

(*C*) While techniques are important in experimental statistics, knowing when to use them and why to use them are more important.

(*D*) In the long run, it does not pay a statistician to fool either himself or his clients.

We have one hypergeneral principle about logical structure, two about statistical method, and one about statisticians. The last may seem to be of smallest scope, but when we consider matters carefully, we see that (*A*), (*B*), and (*C*) all follow from (*D*). To insist on one means or one logical structure for different ends, or to feel that there is *a* solution to the problems of method, are obvious attempts of the statistician to fool himself.

Clearly, one very general consequence is this: "This complexity of experimental statistics will clearly increase."

Reducing the generality somewhat, we list some consequences of (*A*), (*B*), (*C*), and (*D*) which are themselves general principles:

(*A1*) Statistics needs constantly to recognize new ends for which it should try to furnish new means *and* new logical structures.

(*A2*) Statistics needs to avoid over-unification, while encouraging coordination.

(*A3*) Statistical methods should be tailored to the real needs of the user.

(*A4*) Statistics needs continually to compare its own logical structures with the logical structures currently used or being put into use by science, engineering, business, and military administration, and other fields.

(*B1*) In any area of statistical method, analysis cannot be usefully considered alone for more than a limited time; after a time appropriate to the area, design must be brought in.

(*B2*) There are normal sequences (patterns) of growth in immediate ends.

(*B3*) There are normal sequences (patterns) of growth in considerations.

(*B4*) Growth in immediate ends can sometimes be neglected, but growth in considerations is almost never to be neglected.

(*B5*) At any one time, different areas of statistical methodology will be in different states of evolution, both in immediate ends and in considerations.

(*C1*) Competitive statistical techniques indicate a need for manuals of "when to choose which" and not just selection of "the best" technique.

(*C2*) Statisticians owe their clients help in choosing wisely between high confidence in a short inference and low confidence in a long inference.

(*C3*) Techniques of evaluating both the isolated experiment and history down to date will continue to be useful.

(*C4*) "What should be done" is almost always more important than "what can be done exactly." Hence new developments in experimental statistics are more likely to come in the form of approximate methods than in the form of exact ones.

(*D1*) Statisticians must face up to the existence and varying importance of systematic errors.

(*D2*) Statisticians have an obligation to clarify the foundations of their techniques for their clients.

(*D3*) Statisticians should be honest and expository about the relation of precise "assumptions" and exactly "optimum" solutions to real situations.

(*D4*) In every statistical area, we almost certainly need methods admitting one more nuisance parameter, methods of one higher level of robustness and de-parametrization, methods with both of these desiderata.

(*D5*) Statistics must continually study the behavior of its techniques when their conventional assumptions are *not* true.

ILLUSTRATIVE EXAMPLES

I will try to illustrate these principles by discussing particular problems of experimental statistics which show their impact. These examples are not intended to be an exhaustive list. In the light of general principle (*C*), a problem in experimental statistics is not solved by the existence of a mathematical statistical paper showing how to find a solution, or even by the existence of a technique with tables. There is needed an understanding of when and why to use the technique, and this understanding must be spread through a certain minimum number, sometimes small and sometimes large, of experimental statisticians. Thus we may, and should, discuss as unsolved problems some which others may consider as already solved.

(*A1*) *Statistics needs constantly to recognize new ends for which it should try to furnish new means and new logical structures.* A very good illustration of this principle is provided by recent developments in connection with the problem of multiple comparisons. Where one immediate end grew a few years ago, three immediate ends flourish today and promise to flourish for a long time. These three are:

(*1*) The immediate end of providing increments to the store of established knowledge. This to be done by the analysis of existent data with control of the error rate. The analysis to be formulated in confidence or significance statements (*cf.* Tukey [35, 36, 37], Duncan [11, 12, 13] and others).

(*2*) The immediate end of providing protection against too bad a selection among candidates. This to be done by a sequential design of measurement. The result to be selection of the apparently leading candidate when the "stop rule" takes effect. (*cf.* Bechhofer, Dunnett, Sobel [1, 2, 14]).

(*3*) The immediate end of minimizing, in some sense, the sum of the costs of experimentation and the costs of poor choice. This is to be done by a sequential design of measurement. The result to be selection of the apparently leading candidate when the "stop rule" takes effect (*cf.* Grundy, Healy, and Yates [40, 41], Sommerville [31]).

In my judgment, there will be a continuing place for all three immediate ends. To a reasonable extent these places correspond to the terms "basic research," "developmental research," and "operations research," [cp. 22].

This problem of multiple comparisons is still unsolved as a problem of multiple comparisons, because the necessary minimum numbers of experimental statisticians have not yet acquired a working understanding of the new immediate ends involved, or of when which technique is appropriate. Analogous problems, involving immediate ends which differ in analogous ways, are to be expected in more areas of statistics.

(*A2*) *Statistics needs to avoid over-unification, while encouraging co-ordination.* It is now known to mathematical statisticians that all the currently routine modes of statistical technique—significance statements, point estimates, confidence statements, etc.—can be formulated as decision problems. There is a tendency in the air to do so to an increasing degree. This *may* be good mathematical statistics, because it *may* encourage the interchange of useful mathematical techniques among the modes. (We are likely to see in due course whether or not this is true.) But it would surely be very *bad* experimental statistics to treat all these modes in too unified a way. For then some experimental statisticians might be led to forget whether their clients wanted (explicitly or implicitly) a decision or a confidence statement, whether they had done the experiment as a basis for immediate action or as a contribution to knowledge. What more important matter could be forgotten by any experimental statistician?

In almost every area of experimental statistics, there is a problem of providing enough *different* methods to meet the user's needs.

(*A3*) *Statistical methods should be tailored to the real needs of the user.* In a number of cases, statisticians have led themselves astray by choosing a problem which they could solve exactly but which was far from the needs of their clients. They could have chosen a problem closer to their client's needs at the price of an approximate solution. In most of these cases, tailoring the statistical method to the real needs of the client would have meant, and still means, giving up exactness for the sake of usefulness. Realistic assessment of value must urge us to make such "deals" freely and frequently.

The broadest class of such cases comes from the choice of significance procedures rather than confidence procedures. It is often much easier to be "exact" about significance procedures than about confidence procedures. By considering only the most null "null hypothesis" many inconvenient possibilities can be avoided. If the varieties are not different they cannot interact with fertilizers or blocks. If the treatment has no effect, we do not have to be concerned with how its effect varies with the weight or health of the animal or child. And so on—and on. In these examples, it will be clear to many that we are dodging substantial issues.

But throughout experimental statistics there are many areas with significance procedures but without confidence procedures. Almost every one of these areas needs one or more *rough* confidence procedures. Rough procedures will be adequate because the assumptions are not likely to be closely true, so that the probability statements need not follow precisely from the assumptions either. One or more, because techniques based on alternative assumptions give both greater freedom of action and greater confidence in results to the analytical statistician. Here are many unsolved problems in experimental statistics!

At another level of unsolution are the problems where the approximate mathematical statistics has been done, but no use has been made of the results. One outstanding example is the computation by Haldane [19] of the effect of non-normality on the variance of the estimated correlation coefficient. Who has put this to use? Yet it surely is enough to support an empirical robustification procedure involving an effective number of pairs of observations. There must be many more examples like this, where the results have not been carried through to practical usability.

(*A4*) *Statistics needs continually to compare its own logical structures with the logical structures currently used or being put into use by science, engineering, business, and military administration, and other fields.* We can indicate an unsolved problem here which is not likely to be solved in the near future. This is the problem of formalizing some further part of the process of developing new scientific concepts and new scientific theories. Only the most elementary steps in this process have been formalized (in terms of the analysis of conventional types of experiments, of the testing of goodness of fit, and the like). Undoubtedly some, at least, of the less elementary steps can be formalized, but how? And which ones?

This is a vague and diffuse problem, but it is a very important problem indeed. Some would construe it as a problem for philosophers, but I feel that it will require quantitative philosophers (that is, experimental statisticians).

(*B1*) *In any area of statistical method, analysis cannot be usefully considered alone for more than a limited time; after a time appropriate to the area, design must be brought in.* The second and third types of multiple comparison procedures cited above (*A1*) furnish an excellent example of the need for design. For the immediate ends involved the only action, once the measurements are made, is to take the seemingly best candidate. That this is reasonable is, and has been, clear to all. Even a very moderate degree of sophistication was barred from these

situations until the question of when to stop taking measurements was introduced. There must now be many similar cases in other areas today where design considerations have not yet been properly introduced.

(*B2*) *There are normal sequences of growth in immediate ends.* One natural sequence of immediate ends follows the sequence:

(1) Description
(2) Significance statements
(3) Estimation
(4) Confidence statement
(5) Evaluation

In the case of a double binomial the successive levels are illustrative by the sequence of statements.

(*1*) The percentage of success observed among A's was higher than among B's.
(*2*) The percentage of success among A's was significantly greater than among B's.
(*3*) The observed percentage of success among A's exceeded that among B's by a difference of 0.28 in logits. (Or, perhaps, by 15 per cent.)
(*4*) The difference in logits corresponding to the increased percentage of success in A's as against B's is between 0.18 and 0.43 with 95 per cent confidence. (Between 10 per cent and 22 per cent with 95 per cent confidence, perhaps.)
(*5*) Considering both this experiment, and all the observations reported by Smith, Jones, Brown, Robinson, and their coworkers, the indicated difference in logits lies between 0.32 and 0.36 with 5 per cent diffidence (the difference in per cent lies between 17 and 19, perhaps).

The order of (*2*) and (*3*) is not nearly so well defined as that of any other pair. In some areas, and to some experimental statisticians either order would be wrong. We have chosen this order for definiteness and not with sureness.

In the actual case of the double binomial, almost every experimental statistician can handle (*1*), (*2*), and (*3*) easily. Some are not perturbed by (*4*) and of these most but not all can handle (*4*) correctly. No one, so far as the writer knows can treat (*5*) adequately. In other areas we may stop at level (*1*), at level (*2*), at level (*3*), or at level (*4*), but in almost every case there is a next level which represents an unsolved problem.

How to operate at level (*5*) seems to represent an unsolved problem

in many areas. It is a real and important problem, and one whose solution should not be approached flippantly or lightly. Either the classical example of the charge on the electron (as of 1938) or the current example of the heat of sublimation of carbon (which has not improved during the last 25 years) shows that the proper evaluatory answer may be: "The available determinations fall into two systematically different groups, which correspond to values between A and B and between C and D, respectively, and which we are confident cannot be brought into agreement without the introduction of a new systematic adjustment." How many other unusual (from the point of view of formal statistics as found in the books) kinds of conclusions are reasonable in evaluation of all available data? This is not an easy question, but its solution (at least its partial solution) is a prerequisite to that of any problem of evaluation.

There are, of course, other normal sequences of immediate ends, leading mainly through various decision procedures, which are appropriate to development research and to operations research, just as the sequence we have just discussed is appropriate to basic research. (Here "There are, of course" means "There must be! We are sure they exist, but we cannot specify them today.")

(*B3*) *There are normal sequences of growth in considerations.* The area of comparing the typical values of two populations with aid of a sample drawn from each illustrates a customary sequence of evolution in considerations quite nicely. The sequence runs:

(*1*) Normal populations of equal and known variance.
(*2*) Normal populations of general (i.e., probably unequal) and known variances.
(*3*) Normal populations of identical but unknown (but estimated) variance.
(*4*) Normal populations of general and unknown (but estimated) variances.
(*5*) Symmetrical populations of unknown shape and unknown but equal variance.
(*6*) Symmetrical populations of the same unknown shape but general and unknown variances.
(*7*) Symmetrical populations of unknown shapes and variances.
(*8*) Populations of unknown but equal shape and variance.
(*9*) Populations of the same unknown shape and unknown and general variances.
(*10*) Populations of general and unknown shapes and variances.

Here we have exemplified the growth in considerations like these:

(*a*) The scale of the populations might be different.
(*b*) The variance might not be known.
(*c*) The symmetrical populations might not be normal.
(*d*) The populations might not have the same shape.
(*e*) The populations might not be symmetrical.

It is by considering such unpleasant possibilities that we sharpen our techniques and strengthen our understanding.

The normal distribution suffices for levels (*1*) and (*2*), while level (*3*) requires Student's *t*. The next level, (*4*), provides the Fisher-Behrens problem, while (*5*) seems to be the likely end of the direct application of Wilcoxon-Walsh [38–39] procedures (so far only applied to the matched observation case). Beyond this point the *terra* is rather *incognita*, but we may note that through level (*7*) we need to make no distinction between medians and means, while simple rank order procedures are exact through level (*8*).

Not only does this area—and remember that it is one of the most carefully worked over of all areas—provide a good example of a normal sequence of growth in considerations, but it also provides many examples of unsolved problems. The Fisher-Behrens problem arises quite early, at only level (*4*) in the list, yet today the Fisher solution is known not to be unique [33], even in the domain of fiducial probability, while the Aspin-Welch solution may or may not correspond to an exact solution as well as an asymptotic one. What should a poor experimental statistician do?

Who has good-looking solutions for the problems posed by (*5*), (*7*), (*9*), or (*10*)? Who knows how the solutions for level (*4*) just mentioned behave as to error rate when (*5*), (*6*), or (*7*) represents the facts? How do the solutions for level (*4*) behave as to power when either (*4*) or (*3*) represents the facts? And the reader can add many more.

The foreseeable, normal growth in considerations will provide unsolved problems for a long time to come in almost *every area* of statistics.

(*B4*) *Growth in immediate ends can sometmes be neglected, but growth in considerations is almost never to be neglected.* We can use the two-sample area to illustrate this principle also. If we had a clear and reasonable solution to the Fisher-Behrens problem, very few experimental statisticians would dare ignore it. But many are content to teach significance testing without confidence procedures. (The young chemist who can analyze the variance of Latin squares and snatch out single degrees of freedom with zest and ease, but who cannot use Student's *t*

to set confidence limits on $A-B$, because no one ever mentioned it to him, is a poor witness to the teaching of chemists by statisticians!)

(*B5*) *At any one time, different areas of statistical methodology will be in different states of evolution, both in immediate ends and in considerations.* We have only to contrast the two-sample area with the $m \times n$-contingency-table area or the correlation-coefficient area with the measures-of-nonnormality-for-time-series area to find application of this general principle.

(*C1*) *Competitive statistical techniques indicate a need for manuals of "when to choose which" and not just selection of "the best" technique.* Our discussion of the two-sample area should have made it clear that what is needed here is a guide to the various techniques explaining why and when to use them. No selection of a single "best" technique is going to be satisfactory.

Another widely separated area which illustrates the principle nicely is the response maximization area. Here we have a spectrum of suggestions from the carefully thought-out "circle and bee-line (possibly repeated) and then survey" technique of Box and Wilson [5] to the creeping technique of Friedman and Savage [16] and the sophisticated but so far one-dimensional technique of Robbins and Monro [30]. I am sure that all of those named have their place, as do, no doubt, some of the intermediate points in the spectrum. I have, indeed, some idea of where these places are. But I would like to know far more precisely where these places are and why. (You couldn't possibly sell me a *single* best method!)

(*C2*) *Statisticians owe their clients help in choosing wisely between high confidence in a short inference and low confidence in a long inference.* In the analysis of three and more way analyses of variance, there arises the problem of choosing the correct error term (e.g. Goulden [17]). This is the first big problem in the analysis of variance, and one that is still very effective in separating the statisticians from the children. If one classification is years, one choice can be put into words as follows: Will you have differences in average performance averaged over *these particular* years, with narrow confidence limits, or will you have differences in average performance, averaged over a *population of years* of which these years are a sample, with much broader confidence limits. With regard to this particular example, most experimental statisticians are clear and effective. Thus, it may be a solved problem. But in many other areas the corresponding problem is not only unsolved but unposed!

Some have queried the use of "short" and "long" in this context, and

have tried to relate this choice to that of the proper "breadth" of foundation (the advantages of sufficiently broad basis of inference have, of course, been ably discussed by Fisher [15, Section 39]). It is important to avoid possible confusion in this regard. Considerations of breadth arise during the design of an experiment, while considerations of length arise in its interpretation. Thus an experiment to compare certain psychological characteristics within brother-sister pairs would be broadened as to foundation if changed from 50 pairs drawn from Indiana to 5 subgroups of ten pairs each from 5 geographically and culturally separated areas. For *either* experiment, there will be a problem of length of inference! Will we make statements about the average over the 50 pairs of perfectly measured differences, or shall we make statements concerning the average differences in larger populations of which these 50 pairs, or these 5 sets of 10 pairs are a sample or samples? The two questions are quite separate.

(*C3*) *Techniques of evaluating both the isolated experiment and history down to date will continue to be useful.* There are many experimental procedures that involve either the regular measurements of control specimens or the regular use of special calibration procedures. After a new calibration, should we use the old calibration? Should we use only the new calibration? Or should we combine old and new values? With what relative weights? This is a recurrent problem, one whose solution might improve measurement accuracies per dollar in a wide variety of applications. But who has the solution? or better "the solutions," because the path is long from the isolated group of occasional measurements to the production line producing measurements steadily. Different locations along this path will require different solutions. Work on this problem has undoubtedly been hampered by the tradition of the self-contained experiment. But many measurement procedures are far from self-contained experiments.

Like unto this first example is a second. Most procedures of statistical analysis today include a measure of spread in this particular experiment, be it an estimated variance, a total or mean range, or the mean square in a certain line of the analysis of variance. Usually there is past evidence as to the variability in question. In assessing the results of a particular experiment shall we use only the estimate from within the experiment? Only past history? Some combination of the two? Which combination?

This problem of how far to look back is widespread and unsolved. A solution might allow us to narrow the wide confidence limits that go with wide apparent variation and to widen the falsely narrow ones

which go with narrow apparent variation. This would equalize our exposure to error, and tend to let us make sharper statements on the average. Again the philosophy of "each experiment to itself" has stood in the way. But why should we allow this to go on? (Of course the philosophy of "each experiment to itself" is important, of course it must be widely used, but neither always or everywhere! Just another example of (*A2*) and (*C1*).)

(*C4*) *"What should be done" is almost always more important than "what can be done exactly." Hence new developments in experimental statistics are more likely to come in the form of approximate methods than in the form of exact ones.* Once upon a time the calculation of the first four moments was an honorable art in statistics. Then came those who could calculate the exact distributions of simple expressions. And because their results were "exact" they took over the place of honor. (Partly too, perhaps, because the moment calculators failed on occasion to transform their expressions wisely before calculating the moments.) And it came to be *infra dig* to find moments. In seminars one heard A's achievement of calculating the first four moments for n's up to 12 belittled in comparison with B's proof that the distribution tended to normality as n tended to infinity. Yet which result was more useful to the experimental statistician with experimental data for n equal to 5, 10, 20 or even 50—? Probably the first four moments.

If the moments had been on MacArthur's staff, their parting statement would have read "we shall return!" But when? I think that it is high time to bring the calculation of moments back to that high estate which it deserves. We shall always have to deal with messy expressions, whose exact distribution will be found by no one, at least for a long time. Moments may allow us to get on with the work. If they do allow us to do this, let us use them.

The variability of estimates of spectra of time series provides a case in point. Even with the normality assumption, the exact distribution is not going to be easily manageable. Yet the first *two* moments can be found, and found with very useful results. Considerable recent progress in the analysis of physical time series rests on those two moments [e.g. 27, 29].

(*D1*) *Statisticians must face up to the existence and varying importance of systematic errors.* The failure of the statistician to take sufficient cognizance of systematic errors has been in part an escape phenomenon. To a man looking hopefully for a way to shorten a confidence interval by 7 per cent of its length by ingenious devices, the thought of systematic errors which might make it twice as long comes as a severe shock,

and all men try to avoid shocks. Perhaps, too, the recent development of statistics in connection with the uncomfortable sciences like agriculture and biology—uncomfortable because *un*systematic errors tend to be so large—may have much to do with this. Only the sampling survey statisticians, with their recent treatment of "non-sampling errors" seem to be facing up to the existence of systematic errors.

What should experimental statistics as a whole do about systematic errors? Should we change from "95 per cent confidence" to "5 per cent diffidence" and impress on our clients that more diffidence has to be added because of systematic errors? Have we been overselling our clients on the confidence with which they should accept the results of our analyses? Is this why physics is the most-resistant of all the sciences to the penetration of statistics?

Some there will be who will claim that the old ways are good enough, since in comparative experiments the systematic errors tend to be very much smaller than in absolute experiments. Very much smaller, but not zero, is the answer. (The experimental statistician dare not shrink from the war cry of the analyst "Only a fool would use it, but it's better than we used to use!," but on the other hand, he dare not take the motto as a permanent excuse for sloppy methods). Here is a real unsolved problem of experimental statistics; What about systematic errors?

(*D2*) *Statisticians have an obligation to clarify the foundations of their techniques for their clients.* I have the impression that, at the time the analysis of variance was introduced, the practice of adjusting yields for the apparent fertility of blocks was, or would have been, regarded with suspicion—"cooking the observations." Yet the analysis of variance which is quite equivalent in its results, seems to have spread without opposition of this sort. Was this because the arithmetic was so complicated that the poor client didn't understand what was going on? I am sorely afraid that this was the case.

At the beginning, it may have paid the statisticians to fool their clients about the analysis of variance, but does it today? I give vent to a hearty "no!", feeling that many clients get far less out of such analyses than they should, because they don't understand what is going on. How many of your clients really understand what sorts of additive decompositions of the observations underlie the analyses of variance you proudly return to them?

How to explain to the client what the analysis of variance is about? This is surely a problem of experimental statistics. Even if I should know a large part of the answer, as I hope I do, it is an unsolved prob-

lem, since the answer is not at the finger tips of enough experimental statisticians.

In how many other areas are we losing by fooling our clients?

(*D3*) *Statisticians should be honest and expository about the relation of precise "assumptions" and exactly "optimum" solutions to real situations.* As an example here, let us take a field currently under development. Box and his coworkers have been, and continue to be, active in the development of designs for the estimation of all the zeroth, first, and second degree coefficients in a second degree response surface, where the response is a function of 1, 2, 3, 4, 5, etc., variables. In the process he is resting heavily on such "exact" concepts as "orthogonailty" and "estimating all coefficients with the same variance." He is well aware that, because of the way the designs are to be used, these "exact" mathematical properties are not likely to correspond to any physical realities, that, in any particular situation, there is no reason to believe that the "exactly optimum" design is appreciably better than any nearby design. But even if "exactly optimum" does not mean what it says, it may well mean "likely to be quite useful,", as in this case it does.

How many of the potential *users* of such designs will understand that "exactly optimum" doesn't mean what it says? All too few, and for the others we statisticians are likely to be to blame. We have pushed "optimum" procedures for one reason or another, without adequate warning about idealizations and the real world. As a psychologist once said when Mosteller discussed "inefficient statistics" before the Eastern Psychological Association, "inefficient statistics, but efficient statisticians"! How often do we miss the chance to have "non-optimal techniques, but optimal statisticians " apply to us?

Another example of the same sort looms large on the horizon. It concerns all of bioassay and much of the transformation of counted data (a subject about which there are whispers of new discussion). Little attention has been paid to gains or losses from "exact" maximum likelihood, minimum chi-square, or unbiased solutions of bioassay problems. Much attention has been spent in getting these "exact" solutions. Does it matter whether we use logits, probits, or anglits? How much does it matter? (On this there is some information.) What happens if a little non-binomial fluctuation creeps in? Have we been realistic about *anything* in this whole area? Clearly there are many unsolved problems of experimental statistics here.

(*D4*) *In every statistical area, we almost certainly need methods admitting one more nuisance parameter, methods of one higher level of robustness and de-parametrization, methods with both of these desiderata.* Here

we may turn the carpet back to see the dirt—it is a large carpet trying to cover much dirt. We have a reasonably wide variety of procedures for analyzing counted data which assume pure binomial variation. Contingency tables, chi-square, and ω^2 goodness of fit tests, Kolmogoroff-Smirnoff bounds on the population distribution, all-or-none bioassay, and so on. The list is long. Many of the techniques are important. *All of them* need procedures admitting the possibility of additional non-binomial variation. We gave up long ago assuming that we knew the variance of yield of soy bean plots of given size—even though we had empirical data on it. We blithely assume that we know the variance of preparing a dilution and the variance of death among guinea pigs injected with a single dilution—we assume one to be zero and the other to be binomial! We would criticize the varietal trial without an internal estimate of error, yet we look silently on the bioassay without one.

Perhaps in part we have not attacked these problems because of their resemblance to those cited under (*C3*). Perhaps we have not attacked them because their consideration would disturb our clients' techniques or bring to light new sources of variation. But whatever the reasons, they do not seem valid to me today.

Here are many unsolved problems in experimental statistics.

(*D5*) *Statistics must continually study the behavior of its techniques when their conventional assumptions are not true.* I have touched on some minor examples of this principle. Let me cite a few major ones.

Many statistical techniques assume homogeneity of variance, each of them needs a related technique assuming inhomogeneity of variance. How do the present techniques stand up under homogeneity?

Many statistical techniques utilize a normality assumption almost exclusively as a means for predicting the stability of estimated variances. Each needs a related robustified technique which allows for the effects of non-normality on this stability. How do the present techniques stand up under non-normality?

Many discussions of efficiency of estimation assume an underlying normal distribution. Each needs related studies assuming suitably varied nonnormal distributions.

How many unsolved problems do we need?

SOME PROVOCATIVE QUESTIONS

In providing examples of the various general principles, I have indicated a number of unsolved problems of experimental statistics, but there are a few more at the tip of the tongue. In this section I shall seek

to provide a few more, mostly indirectly, by trying to ask some provocative questions.

(*1*) *What are we trying to do with goodness of fit tests?* (Surely not to test whether the model fits exactly, since we know that no model fits *exactly!*) What then? Does it make sense to lump the effects of systematic deviations and over-binomial variation? How should we express the answers of such a test?

(*2*) *Why isn't someone writing a book on one- and two-sample techniques?* (After all, there is a book being written on the straight line!) Why does everyone write another general book? (Even 800 pages is now insufficient for a complete coverage of standard techniques.) How many other areas need independent monograph or book treatment?

(*3*) *Does anyone know when the correlation coefficient is useful, as opposed to when it is used?* If so, why not tell us? What substitutes are better for which purposes?

(*4*) *Why do we test normality?* What do we learn? What should we learn?

(*5*) *How soon are we going to develop a well-informed and consistent body of opinion on the multiple comparison problem?* Can we start soon with the immediate end of adding to knowledge? And even agree on the place of short cuts?

(*6*) *How soon are we going to separate regression situations from comparison situations in the analysis of variance?* When will we clearly distinguish between temperatures and brands, for example, as classifications?

(*7*) *What about regression problems?* Do we help our clients to use regression techniques blindly or wisely? What are the natural areas in regression? What techniques are appropriate in each? How many have considered the "analyses of variance" corresponding to taking out the regression coefficients in *all* possible orders?

(*8*) *What about significance vs. confidence?* How many experimental statisticians are feeding their clients significance procedures when available confidence procedures would be more useful? How many are doing the reverse?

(*9*) *Who has clarified, or can clarify, the problem of nonorthogonal (disproportionate) analysis of variance?* What should we be trying to do in such a situation? What do the available techniques do? Have we allowed the superstition that the individual sums of squares should add up to the total sum of squares to mislead us? Do we need to find new techniques, or to use old ones better?

(*10*) *What of the analysis of covariance?* (There are a few—at least

one [10]—discussions which have been thought about.) How many experimental statisticians know more than one technique of interpretation? How many of these know when to use each? What are all the reasonable immediate aims of using a covariable or covariables? What techniques correspond to each?

(11) *What of the analysis of variance for vectors?* Should we use overt multivariate procedures, or the simpler ones, ones that more closely resemble single variable techniques, which depend on the largest determinantal root? Who has a clear idea of the strength or scope of such methods?

(12) *What of the counting problems of nuclear physics?* (For some of these the physicists have sound asymptotic theory, for others repairs are needed—cf. Link [21].) What happens less asymptotically? What about the use of transformations? What sort of nuisance parameter is appropriate to allow for non-Poisson fluctuations? What about the more complex problems?

(13) *What about the use of transformations?* Have the pros and cons been assembled? Will the swing from significance to confidence increase the use of transformations? How accurate does a transformation need to be? Accurate in doing what?

(14) *Who has consolidated our knowledge about truncated and censored (cf. [18], p. 149) normal distributions so that it is available?* Why not a monograph here that really tells the story? Presumably the techniques and insight here are relatively useful, but how and for what?

(15) *What about range-based methods for more complex situations?* (We have methods for the analysis of single and double classifications based on ranges.) What about methods for more complex designs like balanced incomplete blocks, higher and fractional factorials, lattices, etc.? In which areas would they be quicker and easier? In which areas would they lead to deeper insight?

(16) *Do the recent active discussions about bioassay indicate the solution or impending solution of any problems?* What about logits vs. probits? Minimum chi-square vs. maximum likelihood? Less sophisticated methods vs. all these? Which methods are safe in the hands of an expert? Which in the hands of a novice? Does a prescribed routine with a precise "correct answer" have any value as such?

(17) *What about life testing?* What models should be considered between the exponential distribution and the arbitrary distribution? What about accelerated testing? (Clearly we must use it for long-lived items.) To what extent must we rely on actual service use to teach us about life performance?

(18) *How widely should we use angular randomization* [4]*?* What are its psychological handicaps and advantages? Dare we use it in exploratory experimentation? What will be its repercussions on the selection of spacings?

(19) *How should we seek specified sorts of inhomogeneity of variance about a regression?* What about simple procedures? Can we merely regress the squared deviations from the fitted line on a suitable function? (Let us not depend on normality of distribution in any case!) What other approaches are helpful?

(20) *How soon can we begin to integrate selection theory?* How does the classical theory for an infinite population (as reviewed by Cochran [8]) fit together with the second immediate aim of multiple comparisons (Bechhofer *et al.* [1, 2, 14]) and with the *a priori* views of Berkson [3] and Brown [6]? What are the essential parameters for the characterization of a specific selection problem?

(21) *What are appropriate logical formulations for item analysis (as used in the construction of psychological tests)?* (Surely simple significance tests are inappropriate!) Should we use the method introduced by Eddington [32, pp. 101–4] to estimate the true distribution of selectivity? Should we then calculate the optimum cut off point for this estimated true distribution? Or what?

(22) *What should we do when the items are large and correlated?* (If, for example, we start with 150 measures of personality, and seek to find the few most thoroughly related to a given response or attitude.) What kind of sequential procedure? How much can we rely on routine item analysis techniques? How does experiment for insight differ from experiment for prediction?

(23) *How many experimental statisticians are aware of the problems of astronomy?* What is there in Trumpler and Weaver's book [32] that is new to most experimental statisticians? What in other observational problems like the distribution of nebulae (e.g. [23, 26])?

(24) *How many experimental statisticians are aware of the problems of geology?* What is there in the papers on statistics in geology in the *Journal of Geology* for November 1953 and January 1954 that is new to most experimental statisticians? What untreated problems are suggested there?

(25) *How many experimental statisticians are aware of the problems of meteorology?* What is there in the books of Conrad and Pollak [9] and of Carruthers and Brooks [7] that is new to most experimental statisticians? What untreated problems are suggested there?

(26) *How many experimental statisticians are aware of the problems*

of particle size distributions? What is there in Herdan's book [21] on small particle statistics that is new to most experimental statisticians? What untreated problems are suggested there?

(27) *What is the real situation concerning the efficiency of designs with self-adjustable analyses*—lattices, self-weighted means, etc.—*as compared with their apparent efficiency?* Meier [25] has attacked this problem for some of standard cases, but what are the repercussions? What will happen in other cases? Is there any generally applicable rule of thumb which will make approximate allowance for the biases of unsophisticated procedures?

(28) *How can we bring the common principles of design of experiments into psychometric work?* How can we make allowance for order, practice, transfer of training, and the like through specific designs? Are environmental variations large enough so that factorial studies should always be done simultaneously in a number of geographically separated locations? Don't we really want to factor variance components? If so, why not design psychometric experiments to measure variance components?

(29) *How soon will we appreciate that the columns (or rows) of a contingency table usually have an order?* When there is an order, shouldn't we take this in account in our analyses? How can they be efficient otherwise? Should we test *only* against ordered alternatives? If not, what is a good rule of thumb for allocating error rates? Yates [40] has proposed one technique. What of some others and a comparison of their effectivenesses?

We come now to a set of questions which belong in the list, but which we shall treat only briefly since substantial work is known to be in progress:

(30) What usefully can be done with $m \times n$ contingency tables?

(31) What of a very general treatment of variance components?

(32) What should we really do with complex analyses of variance?

(33) How can we modify means and variances to provide good efficiency for underlying distributions which may or may not be normal?

(34) What about statistical techniques for data about queues, telephone traffic, and other similar stochastic processes?

(35) What are the possibilities of very simple methods of spectral analysis of time series?

(36) What are the variances of cospectral and quadrature spectral estimates in the Gaussian case?

(37) What are useful general representations for higher moments of stationary time series?

Next we revert to open questions:

(38) *How should we measure and analyze data where several coordinates replace the time?* What determines the efficiency of a design? Should we use numerical filtering followed by conventional analysis? How much can we do inside the crater?

(39) *What of an iterative approach to discrimination?* Can Penrose's technique [28] be usefully applied in a multistage or iterative way or both? Does selecting two composites from each of several subgroups and then selecting supercomposities from all these composites pay? If we remove regression on the first two composites from all variables, can we usefully select two new composites from among the residuals?

(40) *Can the Penrose idea be applied usefully to other multiple regression situations?* Can we use either the simple Penrose or the special methods suggested above?

(41) *Is there any sense in seeking a method of "internal discriminant analysis"?* Such a method would resemble factor analysis in resting on no external criterion, but *might* use discriminant-function-like techniques.

(42) *Why is there not a clearer discussion of higher fractionation?* Fractionation (by which we include both fractional factorials and confounding) is reasonably well expounded for the 2^m case. But who can make 3^m, 4^m, 5^m etc. relatively intelligible?

(43) *How many useful fractional factorial designs escape the present group theoretical techniques?* After all, Latin Squares are $k$ths of a k^3, and most transformation sets do not correspond to simple group theory.

(44) *In many applications of higher fractionals, the factors are scaled—why don't we know more about the confounding of the various orthogonal polynomials and their interactions (products)?* Even a little inquiry shows that some particular fractionals are much better than others of the same type.

(45) *What about redundant fractions of mixed factorials?* We know perfectly well that there is no useful simple (nonredundant) fraction of a $2^2 3^3 4^1$, but there may be a redundant one, where we omit some observations in estimating each effect. What would it be like?

A number of further provocative questions have been suggested by others as a result of the distribution of advance copies of this paper and its oral presentation. I indicate some of them in my own words and attitude:

(46) *To what extent should we emphasize the practical power of a test?* Here the practical power is defined as the product of the probability

of reaching a definite decision given that a certain technique is used by the probability of using the technique. (C. Eisenhart)

(47) *What of regression with error in x?* Are the existing techniques satisfactory in the linear case? What of the nonlinear case? (K. A Brownlee)

(48) *What of regression when the errors suffer from unknown auto-correlations?* What techniques can be used? How often is it wise to use them? (K. A. Brownlee)

(49) *How can we make it easier for the statistician to "psychoanalyze" his client?* What are his needs? How can the statistician uncover them? What sort of a book or seminar would help him? (W. H. Kruskal)

(50) *How can statisticians be successful without fooling their clients to some degree?* Isn't their professional-to-client relation like that of a medical man? Must they not follow *some* of the principles? Do statisticians need a paraphrase of the Hippocratic Oath? (W. H. Kruskal)

(51) *How far dare a consultant go when invited?* Once a consultant is trusted in statistical analysis and design, then his opinion is asked on a wider and wider variety of questions. Should he express his opinion on the general direction that a project should follow? Where should he draw the line? (R. L. Anderson)

In closing these questions, it should not be necessary to remind the reader that neither in the last section of examples or in this section of provocative questions have we tried to suggest an order of importance for the unsolved questions suggested. We leave that to the reader.

TOOL BUILDING VS. PROBLEM SOLVING

To judge from published books and articles, experimental statistics has grown by finding tools somehow, and then running around using them. (This impression is undoubtedly *somewhat* inaccurate.) Why has experimental statistics not been more obviously concerned with problems? Partly, perhaps, because it is just beginning to get its growth. Partly, perhaps, because dealing with problems is difficult and likely to lead to approximate solutions. These are valid reasons, but not valid excuses.

As experimental statistics grows toward maturity, it surely should orient more toward areas rather than toward techniques. How much more may be a question. But an essential prerequisite to such reorientation is some picture of what are the areas. This picture will not spring forth full armed, but will come from much work and discussion. As an attempted trigger for this work and discussion, the next section presents a feeble first attempt at classification. Reader, can you do better?

A FEEBLE GUIDE TO AREAS

We shall set up with a digital classification, but without prejudice as to whether the classification provided by one digit is crossed with or nested inside that provided by another. The digits provided will usually not specify an area completely, but they will usually narrow the situation down to a small number of areas.

The first digit classification refers to the general end of the analysis as follows:

(The assessment of, or determination of a wise action in view of)

(*1*) Typical response
(*2*) Variability of response
(*3*) Distribution of response
(*4*) Concealed structures and their coefficients
(*5*) Control charts and other "spotting" procedures
(*9*) Miscellaneous

(If answers are expressible in simple or mixed cumulants, then the degree of these cumulants with respect to response variables is controlling. (*1*) contains cases of degree 1; (*2*) contains cases of degree 2; (*3*) contains cases of higher degree.) Under (*1*) are included regression coefficients as well as means, while correlation analysis considered as a study of predictability comes under (*2*). Contingency tables fall under (*1*), except when the issue is homogeneity, when they fall under (*2*). Factor analysis seems better placed under (*4*) than under (*2*), but structural regression, as practiced in econometrics, seems to fall most naturally under (*1*).

The second digit classification refers to the situation of measurement, and, in description at least, has to be subordinated to the first digit. It runs

(*-1*) Isolated (one or a few) responses, isolated (one or a few) variabilities, isolated (one or a few) distributions, etc.
(*-2*) Response curves or surfaces, variabilities as functions of environmental variables, etc.
(*-3*) Inverse responses (what environment(s) produces a given response), inverse variabilities, etc.
(*-4*) Response to nonenvironmental variable (e.g. time shape of pulses, distribution of grain sizes, power spectrum of time series.)
(*-9*) Miscellaneous

All of bioassay and sensitivity testing will of course be found in (*-3*).

Problems of maximization of response by altering quantitative variables fall best into (*-2*), since attempts to put them into (*-3*) as the search for that environment where the derivatives vanish seem unwise.

The third digit classification refers to the nature of the measurement, and is easy to apply, namely

(*--1*) Absolute measurements without calibration problems
(*--2*) Intermediate cases
(*--3*) Absolute measurements by comparison with a standard
(*--4*) Comparative measurements among a family without calibration problems
(*--5*) Intermediate cases
(*--6*) Comparative measurements among a family with the aid of standards
(*--9*) Miscellaneous

The conventional problems of bioassay fall in (*--3*), while sensitivity to explosion or breakage problems based on falling weights may fall in (*--1*). Conventional comparisons of varieties and fertilizers are usually thought to fall in (*--4*), but must, in many cases, fall in (*--5*).

The fourth digit expresses the kind of response considered, and is again easy to apply. The classes are:

(---, *1*) Directly measured responses
(---, *2*) Responses measured as slopes or regression coefficients
(---, *3*) Adjusted responses (as by covariance)

No examples seem to be needed.

The fifth digit specifies the nature of the response, as follows:

(---, *-1*) Measured response (on reproducible scale)
(---, *-2*) Scored or rated response (by judge or panel)
(---, *-3*) Counted (all-or-none) response
(---, *-9*) Miscellaneous

At the present, the impact of this digit on statistical technique is very noticeable. Should it remain so?

The sixth digit specifies the complexity of the response, as follows:

(---, *--1*) Single variate response
(---, *--2*) Bivariate response
(and so on)
(---, *--8*) Many variate response
(---, *--9*) Miscellaneous

Examples here are not needed.

The seventh digit describes the complexity of the environments considered, as follows:

(---, ---, *1*) Environment varied only randomly
(---, ---, *2*) Environment varied in one measured way
(---, ---, *3*) Environment varied randomly and in one measured way
(---, ---, *4*) Environment varied in two measured ways
(---, ---, *5*) Environment varied in a more complex manner
(---, ---, *9*) Miscellaneous

REFERENCES

[1] Bechhofer, Robert E., Dunnett, Charles W., and Sobel, Milton, "A two-sample multiple decision procedure for ranking means of normal populations with unknown variances," *Annals of Mathematical Statistics*, 24 (1953), 136 (abstract).
[2] Bechhofer, Robert E., and Sobel, Milton, "A sequential multiple decision procedure for ranking means of normal populations with known variances," *Annals of Mathematical Statistics*, 24 (1953), 136 (abstract).
[3] Berkson, Joseph, "'Cost-Utility' as a measure of the efficiency of a test," *Journal American Statistical Association*, 42 (1947), 246–55.
[4] Box, G. E. P., "Multifactor designs of first order," *Biometrika*, 39 (1952), 49–57.
[5] Box, G. E. P., and Wilson, K. B., "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society*, B13 (1951), 1–45.
[6] Brown, George W., "Basic principles for construction and application of discriminators," *Journal of Clinical Psychology*, 6 (1950), 58–60.
[7] Brooks, C. E. P. and Carruthers, N., *Handbook of statistical methods in meteorology*. London, H. M. Stationery Office (1953).
[8] Cochran, W. G., "Improvement by means of selection," *Proceedings of 2nd Berkeley Symposium on Mathematical Statistics and Probability* (1951), 449–70.
[9] Conrad, Victor, and Pollak, L. W., *Methods in Climatology*, 2nd Edition, Cambridge, Mass., Harvard University Press (1950), pp. 459.
[10] DeLury, D. B., "The analysis of covariance," *Biometrics*, 4 (1946), 153–70.
[11] Duncan, David B., "A significance test for differences between ranked treatments in an analysis of variance" *Virginia Journal of Science*, 3 (1951), 172–89 (abstract in *Annals of Mathematical Statistics*, 22 (1951), 142).
[12] Duncan, David B., "On the properties of the multiple comparisons test," *Virginia Journal of Science*, 3 (1952), 49–67.
[13] Duncan, David B., "Multiple range and multiple F tests." Presented to meeting of American Chemical Society, 1953, also Technical Report 6a, Department of Statistics and Statistical Laboratory, Virginia Polytechnic Institute.
[14] Dunnett, Charles W. and Sobel, Milton, "On a multivariate analogue of Student's *t*-distribution, with some tables for the bivariate case." *Annals of Mathematical Statistics*, 24 (1953), 492 (abstract).

[15] Fisher, R. A., *The Design of Experiments*, First Edition, Edinburgh, Oliver and Boyd (1936).

[16] Friedman, Milton, and Savage, L. J., "Planning experiments seeking maxima," Chap. 13 of Statistical Research Group, Columbia University, *Techniques of Statistical Analysis* (edited by Churchill Eisenhart, Millard W. Hastay, and W. Allen Wallis), New York, McGraw-Hill (1947).

[17] Goulden, C. H., *Methods of Statistical Analysis*, 1st Edition, New York, John Wiley and Sons (1939), especially Chapter XI, Section 6 (pp. 122 ff.); 2nd Edition 1952, especially Chapter 5, Section 13 (pp. 90 ff.).

[18] Hald, Anders, *Statistical Theory with Engineering Applications*, New York, John Wiley and Sons (1952).

[19] Haldane, J. B. S., "A note on non-normal correlation," *Biometrika*, 36 (1949), 467–68.

[20] Healy, M. J. R., "Decision between two alternatives—how many experiments," Paper at the third international biometric conference, Bellagio, September, 1953.

[21] Herdan, G., *Small particle statistics*, Amsterdam-Houston-New York-Paris, Elsevier (1953)

[22] Hitchman, Norman, "What is the mission of operations research?" *Journal of the Operations Research Society of America*, 1 (1953), 241–42.

[23] Limber, D. Nelson, "The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field," Submitted to *Astrophysical Journal*.

[24] Link, Richard F., "Some Statistical Techniques Useful for Estimating the Mean Life of a Radioactive Source," Doctoral thesis, Princeton University, 1953.

[25a] Meier, Paul R., "Weighted means and lattice designs," Doctoral thesis, Princeton University.

[25b] Meier, Paul R., "Variance of a weighted mean," *Biometrics*, 9 (1953), 59–73.

[26] Newman, J., Scott, E. L., and Shane, C. D., "On the spatial distribution of the galaxies. A specific model," *Astrophysica Journal*, 117 (1953), 92–138.

[27] Panofsky, H. A., and McCormick, R. A., "The vertical momentum flux at Brookhaven at 109 meters," *Geophysical Research Papers*, 19 (International Symposium Atmos. Turbulence Boundary Layer) (1952), 219–30.

[28] Penrose, L. S., "Some notes on discrimination," *Annals of Eugenics*, 13 (1946–7), 228–37.

[29] Pierson, Willard J., Jr., *A unified mathematical theory for the analysis, propagation and refraction of storm generated ocean surface waves*. Department of Meteorology, New York University, 1952.

[30] Robbins, Herbert, and Monro, Sutton, "A stochastic approximation method," *Annals of Mathematical Statistics*, 22 (1951), 400–7.

[31] Somerville, Paul N., "Optimum sample size for choosing the largest of $k+1$ parameters," Paper at the Institute of Mathematical Statistics meeting Kingston, Ontario, 3 September 1953.

[32] Trumpler, Robert J., and Weaver, Harold F., *Statistical Astronomy*, Berkeley, University of California Press (1953).

[33] Tukey, John W., "Purposes of fiducial inference," Paper before the Institute of Mathematical Statistics, Minneapolis 6 September 1951.

[34] Tukey, John W., "Allowances for various types of error rates," Paper before Institute of Mathematical Statistics, Blacksburg, 19 March 1952.
[35] Tukey, John W., "Multiple Comparisons," Paper before American Statistical Association and Biometric Society, Chicago, 28 December 1952.
[36] Tukey, John W., "Various methods from a unified point of view," Paper before Institute of Mathematical Statistics, Chicago, 29 December 1952.
[37] Tukey, John W., *The problem of multiple comparisons*, In preparation.
[38] Walsh, J. E., "Some significance tests for the median which are valid under very general conditions," *Annals of Mathematical Statistics*, 20 (1949), 64–81.
[39] Wilcoxon, Frank, *Some rapid approximate statistical procedures*, Insecticide and Fungicide Section, Stanford Research Laboratories, American Cyanimide Co., 1948.
[40] Yates, Frank, "The analysis of contingency tables with groupings based on quantitative characters," *Biometrika* 35 (1948), 176–81.
[41] Yates, F., "Principles governing the amount of experimentation in developmental work." *Nature*, 170 (1952), 138–40.

MEASURES OF ASSOCIATION FOR CROSS CLASSIFICATIONS*

Leo A. Goodman and William H. Kruskal
University of Chicago

CONTENTS

* This paper is partly an outgrowth of work sponsored by the Army, Navy, and Air Force through the Joint Services Advisory Committee for Research Groups in Applied Mathematics and Statistics by Contract No. N6ori-02035. We are indebted for helpful comments and criticisms to: Otis D. Duncan (University of Chicago), Churchill Eisenhart (National Bureau of Standards), Maurice G. Kendall

When populations are cross-classified with respect to two or more classifications or polytomies, questions often arise about the degree of association existing between the several polytomies. Most of the traditional measures or indices of association are based upon the standard chi-square statistic or on an assumption of underlying joint normality. In this paper a number of alternative measures are considered, almost all based upon a probabilistic model for activity to which the cross-classification may typically lead. Only the case in which the population is completely known is considered, so no question of sampling or measurement error appears. We hope, however, to publish before long some approximate distributions for sample estimators of the measures we propose, and approximate tests of hypotheses. Our major theme is that the measures of association used by an empirical investigator should not be blindly chosen because of tradition and convention only, although these factors may properly be given some weight, but should be constructed in a manner having operational meaning within the context of the particular problem.

1. INTRODUCTION

MANY studies, particularly in the social sciences, deal with populations of individuals which are thought of as cross-classified by two or more polytomies. For example, the adult individuals living in New York City may be classified as to

Borough:	5 classes
Newspaper most often read:	perhaps 6 classes
Television set in home or not:	2 classes
Level of formal education:	perhaps 5 classes
Age:	perhaps 10 classes

For simplicity we deal largely with the case of two polytomies, although many of our remarks may be extended to a greater number. The double polytomy is the most common, no doubt because of the ease with which it can be tabulated and displayed on the printed page. Most of our remarks suppose the population completely known in regard to the classifications, and indeed this seems to be the way to begin in the construction of rational measures of association. After agreement has been reached on the utility of a measure for a known population, then

(London School of Economics and Political Science), Frederick Mosteller (Harvard University), I. Richard Savage (National Bureau of Standards), Alan Stuart (London School of Economics and Political Science), Louis L. Thurstone (University of North Carolina), John W. Tukey (Princeton University), W. Allen Wallis (University of Chicago), and E. J. Williams (Commonwealth Scientific and Industrial Research Organisation, Australia). Part of Mr. Goodman's work on this paper was carried out at the Statistical Laboratory of the University of Cambridge under a Fulbright Award and a Social Science Research Council Fellowship. The authors were led to work on the problems of this paper as a result of conversations with Louis L. Thurstone and Bernard R. Berelson.

one should consider the sampling problems associated with estimation and tests about this population parameter.

A double polytomy may be represented by a table of the following kind:[1]

A	B				
	B_1	B_2	$\cdots$	B_β	Total
A_1	ρ_{11}	ρ_{12}	$\cdots$	$\rho_{1\beta}$	$\rho_{1\cdot}$
A_2	ρ_{21}	ρ_{22}	$\cdots$	$\rho_{2\beta}$	$\rho_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_α	$\rho_{\alpha 1}$	$\rho_{\alpha 2}$	$\cdots$	$\rho_{\alpha\beta}$	$\rho_{\alpha\cdot}$
Total	$\rho_{\cdot 1}$	$\rho_{\cdot 2}$	$\cdots$	$\rho_{\cdot\beta}$	1

where

Classification A divides the population into the α classes $A_1, A_2, \cdots, A_\alpha$.
Classification B divides the population into the β classes $B_1, B_2, \cdots, B_\beta$.
The proportion of the population that is classified as both A_a and B_b is ρ_{ab}.

The marginal proportions will be denoted by

$\rho_{a\cdot}$ = the proportion of the population classified as A_a.
$\rho_{\cdot b}$ = the proportion of the population classified as B_b.

If the use to which a measure of association were to be put could be precisely stated, there would be little difficulty in defining an appropriate measure. For example, using the above cross-classification of the New York City population, a television service company might wish to

[1] Tables of this kind are frequently called *contingency tables*. We shall not use this term because of its connotation of a specific sampling scheme when the population is not known and one infers on the basis of a sample.

place a single newspaper advertisement which would be read by as many prospective customers as possible. Then the important information from the table of newspaper-most-often-read vs. television-set-in-home-or-not would be: which newspaper is most often read among those with television sets? And a reasonable measure of association would simply be the proportion of those with television sets who read this newspaper.

It is rarely the case, however, that the purpose of an investigation can be so specifically stated. More typically an investigation is exploratory or has a multiplicity of goals. Sometimes a measure of association is desired simply so that a large mass of data may be summarized compactly.

The basic theme of this paper is that, even though a single precise goal for an investigation cannot be specified, it is still possible and desirable to choose a measure of association which has contextual meaning, instead of using as a matter of course one of the traditional measures. In order to choose a measure of association which has meaning we propose the construction of probabilistic models of predictive activity, the particular model to be chosen in the light of the particular investigation at hand. The measure of association will then be a probability, or perhaps some simple function of probabilities, within such a model. Such is our general contention; most of the remainder of this paper is concerned with its exemplification in particular instances.

We wish to emphasize that the specific measures of association described here are *not* presented as factotum or universal measures. Rather, they are suggested as reasonable for use in appropriate circumstances only, and even in those circumstances other measures may and should be considered and investigated.

A good deal of attention has been paid in the literature to the special case of two dichotomies. We are more interested here in measures of association suitable for use with any numbers of classes in the polytomies or classifications.

2. FOUR PRELIMINARY CONSIDERATIONS

Four distinctions or cautionary remarks should be made early in any discussion of measures of association.

2.1. *Continua*

We may or may not wish to think of a polytomy as arising from an underlying continuum. For example, age may for convenience be di-

vided into ten classifications, but it clearly does arise from an underlying continuum; however, newspaper-most-often-read would scarcely be so construed. If a polytomy does arise from an underlying continuum one may or may not wish to assume that the population has some specific kind of distribution with respect to it.

In those cases in which all the polytomies of a study arise jointly from a multivariate normal distribution on an underlying continuum, one would naturally turn to measures of association based on the correlation coefficients. These in turn might well be estimated from a sample by the tetrachoric correlation coefficient method or a generalization of it. In some cases one polytomy may arise from a continuum and the other not. An interesting discussion of this case for two dichotomies was given in 1915 by Greenwood and Yule ([3], Section 3). We do not discuss either of these cases in this paper, but restrict ourselves to situations in which there are no relevant underlying continua.

The desirability of assuming an underlying joint continuum was one of the issues of a heated debate forty years ago between Yule [15] on the one hand and K. Pearson and Heron [9] on the other. Yule's position was that very frequently it is misleading and artificial to assume underlying continua; Pearson and Heron argued that almost always such an assumption is both justified and fruitful.

2.2. *Order*

There may or may not be an underlying order between the classifications of a polytomy. For example "level of formal education" admits an obvious ordering; but borough of residence would not usually be thought of in an ordered way. If there is an ordering, it may or may not be relevant to the investigation. Sometimes an ordering may be important but not its direction. If there is an underlying one-dimensional continuum, it establishes an ordering.

When there is no natural or relevant ordering of the classes of a polytomy, one may reasonably ask that a measure of association not depend on the particular order in which the classes are tabulated.

2.3. *Symmetry*

It may or may not be that one looks at two polytomies symmetrically. When we are sure a priori that a causal relationship (if it exists) runs in one direction but not the other, then our viewpoint will be asymmetric. This will also happen if one plans to *use* the results of the experiment in one direction only. On the other hand, there is often no reason to give one polytomy precedence over another.

2.4. *Manner of Formation of the Classes*

Decisions about the definitions of the classes of a polytomy, or changes from a finer to a coarser classification (or vice-versa), can affect all the measures of association of which we know. For example, suppose we begin with the 4×4 table

0	.25	0	0
.25	0	0	0
0	0	0	.25
0	0	.25	0

and combine neighboring pairs of classes. We obtain

.5	0
0	.5

which might greatly change a measure of association. Or we might combine the three bottom rows and the three right-hand columns. This gives

0	.25
.25	.5

which presents quite a different intuitive degree of association. By other poolings one can obtain other 2×2 tables.

Although this example is extreme, similar changes can be made in the character of almost any cross-classification table. Related examples are discussed by Yule [15].

At first this consideration might seem to vitiate any reasonable discussion of measures of association. We feel, however, that it is in fact desirable that a measure of association reflect the classes *as defined for*

the data. Thus one should not speak, for example, of association between income level and level of formal education without specifying particular class definitions. Of course, in many cases association—however measured—would not be much affected by any reasonable redefinition of the classes, and then the above finicky form of statement can be simplified. That the definition of the classes can affect the degree of association naturally means that careful attention should be given to the class definitions in the light of the expected uses of the final conclusions.

3. CONVENTIONS

It is conventional, and often convenient, to set up a measure of association so that either

(*i*) It takes values between -1 and $+1$ inclusive, is -1 or $+1$ in case of "complete association," and is zero in the case of independence.

(*ii*) It takes values between 0 and $+1$ inclusive, is $+1$ in the case of "complete association," and is zero in the case of independence.

Convention (*i*) is appropriate when the association is thought of as signed (e.g., association between income and dollars spent is positive, between income and per cent of income spent is negative). Convention (*ii*) is appropriate when no such sign considerations exist, as when there is no natural order.

"Complete association," as we shall see, is somewhat ambiguous. "Independence," on the other hand, has its usual meaning, that is

$$(1) \qquad \rho_{ab} = \rho_{a\cdot}\rho_{\cdot b} \; (a = 1, \cdots, \alpha;\, b = 1, \cdots, \beta).$$

Conventions like these have seemed important to some authors, but we believe they diminish in importance as the meaningfulness of the measure of association increases. One real danger connected with such conventions is that the investigator may carry over size preconceptions based upon experience with completely different measures subject to the same conventions. For example, some elementary statistics textbooks warn that a population correlation coefficient less than about .5 in absolute value may have little practical significance, in the sense that then the conditional variance is not much less than the marginal variance. Research workers in various fields thus tend to develop rather strong feelings that population correlation coefficients less than, say, .5, have little substantive importance. The same feelings might be

carried over, without justification, to all other measures of association so defined as to lie between +1 and −1.

It should also be mentioned that once one has a measure of association satisfying one of the above conventions, then an infinite number of others also satisfying the same convention can be obtained—for example, by raising to a power and adjusting the sign if necessary.

4. TRADITIONAL MEASURES

Excellent accounts of these may be found in [16], Chaps. 2 and 3, and [7], Chap. 13. Many of these stem from the standard chi-square statistic upon which a test of independence is usually based. If a finite population has ν members and we set $\nu_{ab}=\nu\rho_{ab}$, $\nu_{a\cdot}=\nu\rho_{a\cdot}$, $\nu_{\cdot b}=\nu\rho_{\cdot b}$, etc., the chi-square statistic in the case of two classifications is

$$(2)\qquad \chi^2 = \sum_a \sum_b \frac{(\nu_{ab} - \nu_{a\cdot}\nu_{\cdot b}/\nu)^2}{\nu_{a\cdot}\nu_{\cdot b}/\nu} = \nu \sum_a \sum_b \frac{(\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2}{\rho_{a\cdot}\rho_{\cdot b}}$$

$$= \nu \sum_a \sum_b \frac{\rho_{ab}^2}{\rho_{a\cdot}\rho_{\cdot b}} - \nu.$$

A great deal of attention has been given to the case $\alpha=\beta=2$. For this special case Yule has defined the following coefficient of association:

$$(3)\qquad Q = \frac{\nu_{11}\nu_{22} - \nu_{12}\nu_{21}}{\nu_{11}\nu_{22} + \nu_{12}\nu_{21}}$$

whose numerator squared is essentially the same as that of a convenient and popular form for χ^2 in the 2×2 case. Another coefficient suggested by Yule for the 2×2 case is

$$(4)\qquad Y = \frac{\sqrt{\nu_{11}\nu_{22}} - \sqrt{\nu_{12}\nu_{21}}}{\sqrt{\nu_{11}\nu_{22}} + \sqrt{\nu_{12}\nu_{21}}}.$$

A coefficient often used for the general $\alpha\times\beta$ case is simply χ^2/ν, often called the mean square contingency and denoted by ϕ^2. A variation of this, suggested by Karl Pearson, is

$$(5)\qquad C = \sqrt{\frac{[\chi^2/\nu]}{1 + \chi^2/\nu}}$$

which has been called the coefficient of contingency, or the coefficient of mean square contingency. Another variation, proposed by Tschuprow, is

$$T = \sqrt{[\chi^2/\nu]/(\alpha-1)(\beta-1)}. \tag{6}$$

The last two suggestions, according to Kendall [7], were made in attempts to norm χ^2 so that it might lie between 0 and 1 and take the extreme values under independence and "complete association." Cramér ([1], p. 282) suggests the following variant:

$$[\chi^2/\nu]/\text{Min}\ (\alpha-1, \beta-1) \tag{7}$$

which gives a better norming than does C or T since it lies between 0 and 1 and actually attains both end points appropriately. Cramér's suggestion does not seem to be well known by workers using this general kind of index.

The fact that an excellent test of independence may be based on χ^2 does not at all mean that χ^2, or some simple function of it, is an appropriate *measure* of degree of association. A discussion of this point is presented by R. A. Fisher ([2], Section 21). We have been unable to find any convincing published defense of χ^2-like statistics as measures of association.

One difficulty with the use of the traditional measures, or of any measures that are not given operational interpretation, is that it is difficult to compare meaningfully their values for two cross-classifications. Suppose that C turns out to be .56 and .24 respectively in two cross-classification tables. One wants to be able to say that there is higher association in the first table than the second, but investigators sometimes restrain themselves, with commendable caution, from making such a comparison. Their restraint may stem in part from the noninterpretability of C. (Of course, when samples are small they may also be restrained by inadequate knowledge of sampling fluctuation.)

One class of measures that will not be discussed here is characterized by the assignment of numerical scores to the classes, followed by the use of the correlation coefficient on these scores. A recent article on such measures is by E. J. Williams [12]. It contains references leading back to earlier literature. We feel that the use of arbitrary scores to motivate measures is infrequently appropriate, but it should be pointed out that measures not motivated by the correlation of scores can often be thought of from the score viewpoint.

5. MEASURES BASED ON OPTIMAL PREDICTION

5.1. *Asymmetrical Optimal Prediction. A Particular Model of Activity*

Let us consider first a probabilistic model which might be useful in a situation of the following kind:

(*i*) Two polytomies, A and B.
(*ii*) No relevant underlying continua.
(*iii*) No natural ordering of interest.
(*iv*) Asymmetry holds: The A classification precedes the B classification chronologically, causally, or otherwise.

An example of such a situation might be a study of the association between college attended (A) and kind of adult occupation (B). Our model of activity is the following: An individual is chosen at random from the population and we are asked to guess his B-class as well as we can, either

1. Given no further information, or
2. Given his A class.

Clearly we can do no worse in case 2 than in case 1. Represent by $\rho_{\cdot m}$ the largest marginal proportion among the B classes and by ρ_{am} the largest proportion in the ath row of the cross-classification table—that is

$$\rho_{\cdot m} = \operatorname*{Max}_b \rho_{\cdot b}, \quad \rho_{am} = \operatorname*{Max}_b \rho_{ab} \, . \tag{8}$$

Then in case 1 we are best off guessing that B_b for which $\rho_{\cdot b} = \rho_{\cdot m}$—that is, guessing that B class which has the largest marginal proportion—and our probability of error is $1 - \rho_{\cdot m}$. In case 2 we are best off guessing that B_b for which $\rho_{ab} = \rho_{am}$ (letting A_a be the given A class)—that is, guessing that B class that has the largest proportion in the observed A class—and our probability of error is[3] $1 - \sum_a \rho_{am}$.

Then we propose as a measure of association (following Guttman [4])

$$\lambda_b = \frac{(\text{Prob. of error in case 1}) - (\text{Prob. of error in case 2})}{(\text{Prob. of error in case 1})} = \frac{\sum_a \rho_{am} - \rho_{\cdot m}}{1 - \rho_{\cdot m}}, \tag{9}$$

which is the relative decrease in probability of error in guessing B_b as between A_a unknown and A_a known. To put this another way, λ_b gives the proportion of errors that can be eliminated by taking account of knowledge of the A classifications of individuals.

Some important properties of λ_b follow:

[3] It may be that in case 1 there is more than one b for which $\rho_{\cdot b} = \rho_{\cdot m}$. Then any method of choosing which of these b's to guess—including flipping an appropriately multi-sided die—gives rise to the same probability of error, $1 - \rho_{\cdot m}$. A similar comment applies to case 2.

(*i*) λ_b is indeterminate if and only if the population lies in one column, that is, lies in one B class.

(*ii*) Otherwise the value of λ_b is between 0 and 1 inclusive.

(*iii*) λ_b is 0 if and only if knowledge of the A classification is of no help in predicting the B classification, i.e., if there exists a b_0 such that $\rho_{ab_0} = \rho_{am}$ for all a.

(*iv*) λ_b is 1 if and only if knowledge of an individual's A class completely specifies his B class, i.e., if each row of the cross-classification table contains at most one nonzero ρ_{ab}.

(*v*) In the case of statistical independence λ_b, when determinate, is zero. The converse need not hold: λ_b may be zero without statistical independence holding.

(*vi*) λ_b is unchanged by permutation of rows or columns.

That λ_b may be zero without statistical independence holding may be considered by some as a disadvantage of this measure. We feel, however, that this is not the case, for λ_b is constructed specifically to measure association in a restricted but definite sense, namely the predictive interpretation given. If there is no association in that sense, even though there is association in other senses, one would want λ_b to be zero. Moreover, all the measures of association of which we know are subject to this kind of criticism in one form or another, and indeed it seems inevitable. To obtain a measure of association one must sharpen the definition of association, and this means that of the many vague intuitive notions of the concept some must be dropped.

We may similarly define

$$\lambda_a = \frac{\sum_b \rho_{mb} - \rho_{m\cdot}}{1 - \rho_{m\cdot}}, \tag{10}$$

where

$$\rho_{m\cdot} = \operatorname*{Max}_a \rho_{a\cdot} \tag{11}$$

$$\rho_{mb} = \operatorname*{Max}_a \rho_{ab}.$$

Thus λ_a is the relative decrease in probability of error in guessing A_a as between B_b unknown and known.

So far as we know, λ_a and λ_b were first suggested by Guttman ([4], Part I, 4), and our development of them is very similar to his.

5.2. *Symmetrical Optimal Prediction. Another Model of Activity*

In many cases the situation is symmetrical, and one may alter the

model of activity as follows: an individual is chosen at random from the population and we are asked to guess his A class half the time (at random) and his B class half the time (at random) either given:

1. No further information, or
2. The class of the individual other than the one being guessed; that is the individual's A_a when we guess B_b and vice versa.

In case 1 the probability of error is $1-\frac{1}{2}(\rho_{\cdot m}+\rho_{m\cdot})$, and in case 2 the probability of error is $1-\frac{1}{2}(\sum_a \rho_{am}+\sum_b \rho_{mb})$. Hence we may consider the relative decrease in probability of error as we go from case 1 to case 2, and define the coefficient

$$\lambda = \frac{\frac{1}{2}[\sum_a \rho_{am} + \sum_b \rho_{mb} - \rho_{\cdot m} - \rho_{m\cdot}]}{1 - \frac{1}{2}(\rho_{\cdot m} + \rho_{m\cdot})}. \tag{12}$$

Some properties of λ follow:

(*i*) λ is determinate except when the entire population lies in a single cell of the table.

(*ii*) Otherwise the value of λ is between 0 and 1 inclusive.

(*iii*) λ is 1 if and only if all the population is concentrated in cells no two of which are in the same row or column.

(*iv*) λ is 0 in the case of statistical independence, but the converse need not hold.

(*v*) λ is unchanged by permutations of rows or columns.

(*vi*) λ lies between λ_a and λ_b inclusive.

The computation of λ_a, λ_b, or λ is extremely simple. Usually one is given the population, not in terms of the ρ_{ab}'s but rather in terms of the numbers of individuals in each cell. Let ν be the total number of individuals in the population, $\nu_{ab}=\nu\rho_{ab}$, $\nu_{am}=\nu\rho_{am}$, $\nu_{mb}=\nu\rho_{mb}$, and so on. Then

$$\lambda_b = \frac{\sum_a \nu_{am} - \nu_{\cdot m}}{\nu - \nu_{\cdot m}}, \tag{13}$$

$$\lambda_a = \frac{\sum_b \nu_{mb} - \nu_{m\cdot}}{\nu - \nu_{m\cdot}}, \tag{14}$$

$$\lambda = \frac{\sum_a \nu_{am} + \sum_b \nu_{mb} - \nu_{\cdot m} - \nu_{m\cdot}}{2\nu - (\nu_{\cdot m} + \nu_{m\cdot})}. \tag{15}$$

5.3. *An example*

The following table is taken from reference [7], p. 300, and originally was given by Ammon in "Zur Anthropologie der Badener." It deals with hair and eye color of males. The table is given in terms of the ν_{ab}'s. A_1, A_2, A_3 are respectively Blue, Grey or Green, Brown; B_1, B_2, B_3, B_4 are respectively Fair, Brown, Black, Red.

Eye Color Group	Hair Color Group				
	B_1	B_2	B_3	B_4	$\nu_{a\cdot}$
A_1	1768	807	189	47	2811
A_2	946	1387	746	53	3132
A_3	115	438	288	16	857
$\nu_{\cdot b}$	2829	2632	1223	116	$\nu=6800$

We have:

$$\nu_{1m} = 1768 \qquad \nu_{m1} = 1768$$
$$\nu_{2m} = 1387 \qquad \nu_{m2} = 1387$$
$$\nu_{3m} = 438 \qquad \nu_{m3} = 746$$
$$\nu_{m4} = 53$$
$$\nu_{\cdot m} = 2829 \qquad \nu_{m\cdot} = 3132$$

$$\lambda_a = \frac{3,954 - 3,132}{6,800 - 3,132} = \frac{822}{3,668} = .2241$$

$$\lambda_b = \frac{3,593 - 2,829}{6,800 - 2,829} = \frac{764}{3,971} = .1924$$

$$\lambda = \frac{822 + 764}{3,668 + 3,971} = \frac{1,586}{7,639} = .2076.$$

(Quotients are given to four places.) The traditional measures of association have the following values: $\chi^2/\nu = .1581$, $C = .3695$, $T = .2541$, Cramér's measure $= .2812$.

This example appears as an illustration of the usual approach to measures of association in [7], a standard statistical reference work. It is not hard to think of interpretations or variations in which one

of the λ coefficients would be appropriate. For example, one might be studying the efficacy of an identification scheme for males in which hair color was given but not eye color. Another example might be in connection with a study of popular beliefs about the relationship between hair color and eye color.

5.4. *Weighting Columns or Rows*

In some cases, particularly when comparisons between different populations are important, the measures λ_a, λ_b, or λ may not be suitable, since they depend essentially on the marginal frequencies. To put this in terms of the model of activity: in some cases we do not want to think of choosing an individual from the actual population *at hand* in a random way, but rather from some other population which is related to the actual population in terms of conditional frequencies.

This point is stressed by Yule in reference [15] and is illustrated by the kind of medical example[3] given there. Suppose that we are concerned with the effects of a medical treatment on persons contracting an often fatal disease. Very large samples from two different hospitals are available, giving the following ρ_{ab} tables:

	Hospital I			Hospital II		
	Lived	Died	Total	Lived	Died	Total
Treated	.84	.04	.88	.42	.02	.44
Not treated	.03	.09	.12	.14	.42	.56
Total	.87	.13	1.00	.56	.44	1.00

Here the A classes are Treated or Not-treated, and the B classes Lived or Died. The given numbers are ρ's and marginal ρ's.

We are interested in the association between treatment and life, and might conclude that λ_b would be an appropriate measure of this. We find

$$\lambda_b \text{ for Hospital I} = \frac{.93 - .87}{.13} = .462$$

$$\lambda_b \text{ for Hospital II} = \frac{.84 - .56}{.44} = .636.$$

[3] We do not wish to suggest by this example that λ_b is necessarily appropriate as a measure of association between treatment and cure. A very interesting discussion of this medical case has been given by Greenwood and Yule [3] who bring out many difficulties and suggest various viewpoints. Another interesting paper on the medical 2×2 table is that of Youden [14].

Yet the *conditional* probabilities of life, given treatment (nontreatment), are exactly the same for both hospitals, namely .955 (.250). The reason that the conditional probabilities are the same while the λ_b values are different is, of course, that the two hospitals treated very different proportions of their patients. And the proportions treated were probably determined by factors having nothing to do with 'inherent' association between treatment and cure.

It may seem reasonable in such a case as this to replace our model of activity by one in which an individual is drawn from the population so that the probability of his being in any given A_a is exactly $1/\alpha$, i.e., so that all A classes are equiprobable; and with conditional B class probabilities equal to those of the original population. That is to say, it may seem reasonable to replace the quantities ρ_{ab} by the quantities

$$\frac{1}{\alpha}\frac{\rho_{ab}}{\rho_{a\cdot}} \tag{16}$$

and use this as the population to which λ_b is applied. We may thus define, in terms of the conditional probabilities given A_a,

$$\lambda_b^* = \frac{\frac{1}{\alpha}\sum_a \frac{\rho_{am}}{\rho_{a\cdot}} - \frac{1}{\alpha}\operatorname*{Max}_b \sum_a \frac{\rho_{ab}}{\rho_{a\cdot}}}{1 - \frac{1}{\alpha}\operatorname*{Max}_b \sum_a \frac{\rho_{ab}}{\rho_{a\cdot}}}. \tag{17}$$

If we do this in the present example, we get, of course, the same altered ρ table for both hospitals

.477	.023	.500
.125	.375	.500
.602	.398	1.00

and in both cases

$$\lambda_b^* = \frac{.250}{.398} = .628.$$

An analogous procedure could be used to define λ_a^* and λ^*. Note also
measu
It is n

that other 'artificial' marginal ρ's besides .5 could be used if appropriate. Yule [15] suggests as a desideratum for coefficients of association their invariance under transformations on the $\{\rho_{ab}\}$ matrix of form

$$\rho_{ab} \to s_a t_b \rho_{ab},\ s_a,\ t_b > 0; \quad a = 1, \cdots, \alpha; \quad b = 1, \cdots, \beta.$$

Such a transformation may readily be found (at least when no $\rho_{ab}=0$) to make *all* four marginals of a two by two table equal to .5. In this connection, we refer to a recent article by Pompilj [10] in which such transformations are carefully discussed.

All further measures may be considered for unweighted or weighted marginal proportions, whichever are appropriate.

6. MEASURES BASED UPON OPTIMAL PREDICTION OF ORDER

6.1. *Preliminaries*

Heretofore we have considered measures of association suitable for the unordered case, that is, measures which do not change if the columns (rows) are permuted. Now we shall suggest a measure suitable for the ordered case. Suppose that the situation is of the following kind:

(*i*) Two polytomies, A and B.
(*ii*) No relevant underlying continua.
(*iii*) Directed ordering is of interest.
(*iv*) The two polytomies appear symmetrically.

By (*iii*) we mean that we wish to distinguish, in the 3×3 case between, for example,

ρ_{11}	0	0
0	ρ_{22}	0
0	0	ρ_{33}

and

0	0	ρ_{13}
0	ρ_{22}	0
ρ_{31}	0	0

calling the first of these complete association and the second complete counterassociation. We may wish to make the convention that in these two cases the proposed measure should take the values +1 and −1 respectively. If the sense or direction of order is irrelevant we can, for example, simply take the absolute value of a measure appropriate to directed ordering.

There are vaguenesses in the idea of complete ordered association. For example, everyone would probably agree that the following case is one of complete association:

0	0	0
ρ_{21}	0	0
0	ρ_{32}	0

The following situation is not so clear:

ρ_{11}	0	0
ρ_{21}	ρ_{22}	0
0	ρ_{32}	ρ_{33}

As before, the procedure we shall adopt toward this and toward more complex questions is to base the measure of association on a probabilistic model of activity which often may be appropriate and typical.

6.2. *A Proposed Measure*

Our proposed model will now be described. Suppose that two individuals are taken independently and at random from the population (technically with replacement, but this is unimportant for large populations). Each falls into some (A_a, B_b) cell. Let us say that the first falls in the $(A_{\underline{a}_1}, B_{\underline{b}_1})$ cell, and the second in the $(A_{\underline{a}_2}, B_{\underline{b}_2})$ cell. (Underlined letters denote random variables.) $\underline{a}_i$ $(i=1, 2)$ takes values from 1 to α; $\underline{b}_i$ $(i=1, 2)$ takes values from 1 to β.

If there is independence, one expects that the order of the $\underline{a}$'s has no connection with the order of the $\underline{b}$'s. If there is high association one expects that the order of the $\underline{a}$'s would generally be the same as that of the $\underline{b}$'s. If there is high counterassociation one expects that the orders would generally be different.

Let us therefore ask about the probabilities for like and unlike or-

ders. In order to avoid ambiguity, these probabilities will be taken conditionally on the absence of ties. Set

$$\text{(18)}\quad \Pi_s = \Pr\{\underline{a}_1 < \underline{a}_2 \ \text{ and } \ \underline{b}_1 < \underline{b}_2; \ \text{ or } \ \underline{a}_1 > \underline{a}_2 \ \text{ and } \ \underline{b}_1 > \underline{b}_2\}$$

$$\text{(19)}\quad \Pi_d = \Pr\{\underline{a}_1 < \underline{a}_2 \ \text{ and } \ \underline{b}_1 > \underline{b}_2; \ \text{ or } \ \underline{a}_1 > \underline{a}_2 \ \text{ and } \ \underline{b}_1 < \underline{b}_2\}$$

$$\text{(20)}\quad \Pi_t = \Pr\{\underline{a}_1 = \underline{a}_2 \ \text{ or } \ \underline{b}_1 = \underline{b}_2\}.$$

Then the conditional probability of like orders given no ties is $\Pi_s/(1-\Pi_t)$ and the conditional probability of unlike orders given no ties is $\Pi_d/(1-\Pi_t)$. Of course, the sum of these two quantities is one.

A possible measure of association would then be $\Pi_s/(1-\Pi_t)$, but it is a bit more convenient to look at the following quantity:

$$\text{(21)}\qquad \gamma = \frac{\Pi_s - \Pi_d}{1 - \Pi_t}$$

or the *difference* between the conditional probabilities of like and unlike orders. In other words γ tells us how much more probable it is to get like than unlike orders in the two classifications, when two individuals are chosen at random from the population.

Since $\Pi_s + \Pi_d = 1 - \Pi_t$, we may write γ as

$$\text{(22)}\qquad \gamma = \frac{2\Pi_s - 1 + \Pi_t}{1 - \Pi_t}$$

which is convenient for computation, using the easily checked relationships

$$\text{(23)}\qquad \Pi_s = 2\sum_a \sum_b \rho_{ab}\Big\{\sum_{a'>a}\sum_{b'>b} \rho_{a'b'}\Big\}$$

$$\text{(24)}\qquad \Pi_t = \sum_a \rho_{a\cdot}^2 + \sum_b \rho_{\cdot b}^2 - \sum_a \sum_b \rho_{ab}^2.$$

Some important properties of γ follow:

(*i*) γ is indeterminate if the population is concentrated in a single row or column of the cross-classification table.

(*ii*) γ is 1 if the population is concentrated in an upper-left to lower-right diagonal of the cross-classification table. γ is -1 if the population is concentrated in a lower-left to upper-right diagonal of the table.

(*iii*) γ is 0 in the case of independence, but the converse need not hold except in the 2×2 case. An example of nonindependence with $\gamma=0$ is

.2	0	.2
0	.2	0
.2	0	.2

For tables up to 5×5 with ρ's expressed to two decimal places computation is fairly rapid. If many tables of the same size are at hand a cardboard template would be convenient. A check on Π_s is to recompute using inverted ordering in both dimensions. γ may be rewritten in terms of the ν's by putting "ν_{ab}" for "ρ_{ab}," etc., and replacing "1" in (22) by "ν^2."

In the 2×2 case we find that

$$\gamma = \frac{\rho_{11}\rho_{22} - \rho_{12}\rho_{21}}{\rho_{11}\rho_{22} + \rho_{12}\rho_{21}}. \tag{25}$$

This is the same as Yule's coefficient of association Q mentioned in Section 4. In this case $\gamma = \pm 1$ if any one cell is empty. For example,

ρ_{11}	0
ρ_{21}	ρ_{22}

gives rise to $\gamma = 1$ always.

Any case of the following forms will give rise to $\gamma = 1$, since a conflict in order is impossible:

ρ_{11}	ρ_{12}	0
0	ρ_{22}	ρ_{23}
0	0	ρ_{33}

ρ_{11}	0	0
ρ_{21}	0	0
ρ_{31}	ρ_{32}	ρ_{33}

The right-hand table might be thought of as a case of "complete curvilinear association."

Stuart [11], starting from a suggestion by Kendall [6], has proposed a measure of association in the ordered case much like γ. Stuart's measure, which he calls τ_c is, in our notation,

$$\tau_c = \frac{\Pi_s - \Pi_d}{(m-1)/m}$$

where $m = \text{Min}(\alpha, \beta)$. The term $(m-1)/m$ is introduced in order that τ_c may attain, or nearly attain, the absolute value 1 when the entire population lies in a longest diagonal of the table. Stuart develops his measure by considering a two-way ordered classification table as two rankings of a population, where many ties appear in one or both rankings as two individuals of the population fall in the same column or row or both. Then each ordered pair of individuals is assigned a score with respect to each ranking: 0 if there is a tie, or ± 1 as one or the other is ranked higher. Finally the product-moment correlation coefficient is formally computed with these scores, and the norming factor is introduced.

Thus, our development of γ is seen to give another and more natural interpretation for the numerator of τ_c: it is the probability of like order less the probability of unlike order when two individuals are chosen at random. In addition the form in which τ_c is given above, together with (23) and (24), suggests a computation procedure somewhat different than that of [11].

6.3. *An Example*

Whelpton, Kaiser, and others [17] have investigated in great detail relationships between human fertility and a number of social and psychological characteristics of married couples. The analyses resulting from these investigations are replete with cross-classification tables, together with accompanying verbal explanations and recapitulations. Numerical indexes of association appear to have been used rarely, if at all, in this work.

We wish to examine briefly one of these cross-classification tables as an example of a cross-classification with an order in both classifications. This examination should be construed neither as approval nor criticism of the methodology used in the studies edited by Whelpton and Kaiser, for this would not be appropriate here. (The reader may refer to [18] and [19] for critical reviews.) However, we do feel that the use of summarizing indexes of association in a study of this kind may well be worth while for at least two reasons. One is that the reader finds it very difficult to obtain a bird's-eye view of the extensive numerical material without depending almost wholly on the author's own conclusions. Second, the use of indexes would mitigate the criticism that the author, consciously or not, selects from his numerical data

those comparisons that are in line with his a priori beliefs. Needless to say, an index of association is recommended by these arguments only if it has some reasonable interpretation.

The particular table we wish to consider follows, in terms of numbers of married couples. It refers to a rather special, but well defined, population: white Protestant married couples living in Indianapolis, married in 1927, 1928, or 1929, and so on. The data were obtained by stratified sampling, with strata based on numbers of live births. However, for present purposes we do not consider any questions of sampling, response error, specification of population, etc. The table is condensed from a more detailed cross-classification given in [17], vol. 2, pp. 286, 389, and 402. Further, we shall not define the fertility-planning categories that follow, but merely indicate the order.

CROSS-CLASSIFICATION BETWEEN EDUCATIONAL LEVEL OF WIFE AND FERTILITY-PLANNING STATUS OF COUPLE. SOURCE [17], VOL. 2. NUMBERS IN BODY OF TABLE ARE FREQUENCIES

Highest level of formal education of wife	Fertility-planning status of couple				Row totals
	A Most effective planning of number and spacing of children	B	C	D Least effective planning of children	
one year college or more	102	35	68	34	239
3 or 4 years high school	191	80	215	122	608
less than 3 years high school	110	90	168	223	591
Column totals	403	205	451	379	1438

This is clearly a case where there is relevant order in both classifications. We may first compute Π, as follows (schematically):

$$\Pi_s = \frac{2}{(1438)^2}[102(80 + 90 + 215 + 168 + 122 + 223)$$
$$+ 35(215 + 168 + 122 + 223) + \cdots + 215(223)]$$
$$= \frac{2}{(1438)^2}[102 \times 898 + 35 \times 728 + \cdots + 215 \times 223]$$
$$= \frac{2 \times 311{,}632}{2{,}067{,}844} = .301.$$

This means that if we pick two couples at random from those included in the table, the probability is .301 that they are not tied in either classification and that they fall in the same order for both classifications (e.g., if educational level of wife is greater for first couple chosen, then effectiveness of fertility planning is also greater).

Similarly we compute that $\Pi_d = .163$. This is the probability of no ties and *different* orders. Finally Π_t, the probability of a tie in at least one classification, is .536. Note that $\Pi_s + \Pi_d + \Pi_t = 1.000$.

The conditional probability of like order, given no tie, is $\Pi_s/(1 - \Pi_t) = .301/.464 = .649$; and the conditional probability of unlike order is $.163/.464 = .351$. Clearly there is a greater chance of like order than of unlike order, and this means positive association, if the operational model is a reasonable one. To measure the magnitude of this association we may use γ, which here is equal to

$$\frac{.301 - .163}{.464} = .298.$$

This is the difference between the conditional probabilities of like and unlike order, given no ties.

It might be thought that one should look, not at the actual population above, but at a related population with equal row totals and with the same relative frequencies within each row. That is, we might wish to work with a derived population within which one-third of the wives lie in each education category, but which is otherwise the same. This derived population is readily obtained (in terms of its ρ_{ab}'s) by dividing each frequency in the above table by three times the total in its row. Very minor adjustments were made because of rounding, in order that the over-all sum be 1.000. For the same reason, the row totals are not exactly equal.

CROSS-CLASSIFICATION BETWEEN EDUCATIONAL LEVEL OF WIFE AND FERTILITY-PLANNING STATUS OF COUPLE. DERIVED FROM PRIOR TABLE BY ADJUSTMENT TO MAKE ROW TOTALS EQUAL. NUMBERS IN BODY OF TABLE ARE RELATIVE FREQUENCIES (ρ_{ab}'s).

Highest level of formal education of wife	Fertility-planning status of couple				Row totals
	A Most effective planning of number and spacing of children	B	C	D Least effective planning of children	
one year college or more	.142	.049	.095	.047	.333
3 or 4 years high school	.105	.044	.118	.067	.334
less than 3 years high school	.062	.050	.095	.126	.333
Column totals	.309	.143	.308	.240	1.000

For this table we find $\Pi_s = .325$, $\Pi_d = .170$, $\Pi_t = .505$.

Hence $\Pi_s/(1-\Pi_t) = .657$, $\Pi_d/(1-\Pi_t) = .343$, and $\gamma = .314$. There is no great difference between the original and the adjusted table in regard to association as measured by probabilities of like and unlike order.

Alternatively, one might wish to adjust the tabular entries so that column totals are equal, or one might attempt to adjust the entries so that the row totals are equal and the column entries are equal.

7. THE GENERATION OF MEASURES BY THE INTRODUCTION OF LOSS FUNCTIONS

7.1. *Models Based on Loss Functions*

Instead of obtaining a measure as a natural function of probabilities in the context of a model of predictive behavior, one can more generally employ loss functions. In such a way, one can even artificially generate the conventional measures described in Section 4.

7.2. *Loss Functions and the λ Measures*

In the context of Section 5.1 let us suppose that in guessing an individual's B class one incurs a loss $L(b_1, b_2)$, where B_{b_1} is the true B class and B_{b_2} is the guessed one. Consider first guessing B_b given no information. Then a scheme of guessing B_b with probability $p_b(p_b \geqq 0, \sum p_b = 1)$ leads to an average loss of $\sum_{b_1} \sum_{b_2} \rho_{\cdot b_1} p_{b_2} L(b_1, b_2)$. It is easily seen that this average is minimized by guessing that B_{b_2} for which $\sum_b \rho_{\cdot b} L(b, b_2)$ is a minimum, or if there are two or more minima by guessing any one of them. Let b_L be any one of these b_2's, so that the minimum average loss is $\sum_b \rho_{\cdot b} L(b, b_L)$.

On the other hand if the individual's A class is known to be A_a, the best scheme of guessing is to select b_2 to minimize $\sum_b \rho_{ab} L(b, b_2)$. Let b_{La} be such a minimizing b_2; then the minimum average loss when A_a is known is $\sum_b (\rho_{ab}/\rho_{a\cdot}) L(b, b_{La})$, and the over-all minimum average loss with A_a's known is $\sum_a \sum_b \rho_{ab} L(b, b_{La})$.

The decrease in loss as we pass from the first case to the second is therefore

$$\sum_b \rho_{\cdot b} L(b, b_L) - \sum_a \sum_b \rho_{ab} L(b, b_{La}). \tag{26}$$

It would be reasonable to norm this by division by the first term, $\sum_b \rho_{\cdot b} L(b, b_L)$, to obtain a generalization of λ_b.

Notice that if $L(b_1, b_2)$ is 0 when $b_1 = b_2$ and 1 when $b_1 \neq b_2$, we obtain exactly λ_b. Analogous procedures give us generalizations of λ_a and λ. A slight extension of the procedure, permitting the loss to depend on the true A class as well as the true and guessed B classes, gives a generalization of λ_b^*.

7.3. *The Conventional Measures in Terms of Loss Functions*

Suppose, instead of predicting the classes of individuals, we are asked to determine the values ρ_{ab} when only the $\rho_{a\cdot}$ and $\rho_{\cdot b}$ are known. In the case of independence, these ρ_{ab} are $\rho_{a\cdot}\ \rho_{\cdot b}$. In the more general case, the difference between ρ_{ab} and $\rho_{a\cdot}\ \rho_{\cdot b}$ may be thought of as the amount of error made by assuming independence. If the loss is proportional to the square of the error, inversely proportional to the estimate $\rho_{a\cdot}\ \rho_{\cdot b}$, and additive, we have

$$\sum_a \sum_b k_{ab} \frac{(\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2}{\rho_{a\cdot}\rho_{\cdot b}} \tag{27}$$

where the k_{ab}'s are given constants. For comparison with standard chi-square, express this in terms of the ν_{ab}'s

$$\sum_a \sum_b k_{ab} \frac{\left(\nu_{ab} - \dfrac{\nu_{a\cdot}\nu_{\cdot b}}{\nu}\right)^2}{\nu_{a\cdot}\nu_{\cdot b}} \tag{28}$$

and finally set $k_{ab} = \nu$ to obtain just the chi-square statistic.

Although this procedure and loss function seem to us rather artificial, they do give one way of motivating the chi-square statistic as a measure of association.

8. RELIABILITY MODELS

8.1. *Generalities*

Consider now cases in which the classes are the *same* for the two polytomies, so that we deal with an $\alpha \times \alpha$ table, but differ in that assignment to class depends on which of two methods of assignment is used. Thus we might for example consider two psychological tests both of which classify deranged individuals as to the type of mental disorder from which they suffer. Or again, we might consider two observers taking part in a sociological experiment wherein they independently and subjectively rate each child in a group of children on a five point scale for degree of cooperation.

One is often concerned in such cases with the degree to which the two methods of assignment to class agree with each other. In the case of the psychological tests, for example, one of the tests might be a well established standard procedure and the other might be a more easily applied variant under consideration as a substitute. The psychologist would probably only consider the variant seriously if it gave the same answers as the standard test often enough in some sense which he would have to explicate. In the case of the two observers, the problem might be whether the kind of subjective ratings given by trained observers in that context are similar enough to warrant the use of such subjective ratings at all.

As before we shall not consider here sampling problems, but rather shall suppose the population ρ_{ab}'s known. The several distinctions and conventions of Sections 2 and 3 apply here of course, but the measures suggested in Sections 5 and 6 do not seem appropriate in this reliability

context. One reason is that the classes are the same for both polytomies. This means that even in the unordered case we do *not* want a measure which is invariant under interchange of rows and interchange of columns unless the two interchanges are the same.

An obvious measure of reliability in such a study is just $\sum_a \rho_{aa}$, the probability of agreement. However, we shall also consider some other possibilities.

8.2. *A Measure of Reliability in the Unordered Case*

The measure we shall now propose might be appropriate under the following conditions:

(*i*) Two polytomies are the same, but arise from different methods of assignment to class.
(*ii*) No relevant underlying continua.
(*iii*) No relevant ordering.
(*iv*) Our interest in reliability is symmetrical as between the two polytomies.

A modal class over both classifications is any $A_a(=B_a)$ such that $\rho_{a\cdot}+\rho_{\cdot a} \geqq \rho_{a'\cdot}+\rho_{\cdot a'}$ for all a'. It is simplest to suppose that there is a unique modal class, but if there are more any can be chosen. Denote by $\rho_{M\cdot}$ and $\rho_{\cdot M}$ the two marginal proportions corresponding to the modal class.

A modal class can be given the following interpretation: choose an individual at random from the population and pick one of the two methods of assignment by flipping a fair coin. What is the long-run best guess beforehand of how the chosen method will classify the chosen individual? The answer is: the modal class; and if the modal class is A_a, then the probability of a correct guess is $\frac{1}{2}(\rho_{a\cdot}+\rho_{\cdot a}) = \frac{1}{2}(\rho_{M\cdot}+\rho_{\cdot M})$.

In so far as there is good reliability between the two methods of assignment, one could make a better guess if one knew how the other method of assignment would classify the individual, and then followed the rule of guessing the *same* class for the method being predicted. The probability of a correct guess would then be $\sum \rho_{aa}$. Thus as we go from the no information situation to the other-method-known situation, the probability of error decreases by $\sum \rho_{aa} - \frac{1}{2}(\rho_{M\cdot}+\rho_{\cdot M})$. This quantity may vary from $-\frac{1}{2}$ to $1-(1/\alpha)$. It takes the value $-\frac{1}{2}$ when all the diagonal ρ_{aa}'s are zero and the modal probability, $\rho_{M\cdot}+\rho_{\cdot M}$ is 1. It takes the value $1-(1/\alpha)$ when the two methods always agree and each category is equi-probable.

To get a measure we should alter the above quantity, since a sufficiently large ρ_{aa} for some a will make the above quantity low even though $\sum \rho_{aa}$ is nearly 1. It seems reasonable to norm by division by the probability of error given no information, that is by $1-\frac{1}{2}(\rho_{M\cdot}+\rho_{\cdot M})$. Hence we propose the measure

$$\lambda_r = \frac{\sum \rho_{aa} - \frac{1}{2}(\rho_{M\cdot} + \rho_{\cdot M})}{1 - \frac{1}{2}(\rho_{M\cdot} + \rho_{\cdot M})}. \tag{29}$$

This may be interpreted as the relative decrease in error probability as we go from the no information situation to the other-method-known situation.

The measure λ_r can take values from -1 to 1. It takes the value -1 when all the diagonal ρ_{aa}'s are zero and the modal probability, $\rho_{M\cdot}+\rho_{\cdot M}$ is 1. It takes the value 1 when the two methods always agree. λ_r is indeterminate only when both methods always give only one and the same class. In the case of independence λ_r assumes no particular value. This characteristic might be considered a disadvantage, but it seems to us that an index of this kind would only be used where there is known to be dependence between the methods, so that misbehavior of the index for independence is not important.

8.3. *Reliability in the Ordered Case*

For the case in which the classes are ordered, but a meaningful metric is absent, we have been unable to find a measure better than one of the following kind:

$$\sum_{a=1}^{\alpha} \rho_{aa} \quad \text{(as suggested in Section 8.1)} \tag{30a}$$

$$\sum_{|a-b|\leq 1} \rho_{ab} \tag{30b}$$

$$\sum_{|a-b|\leq 2} \rho_{ab}, \tag{30c}$$

that is, the only reasonable measures we know of are those that are based upon either the probability of agreement, the probability of agreement to within one neighboring class, two neighboring classes, and so on. If desired one could weight these probabilities when classification in a neighboring class is not as desirable as in the same class. Thus one might consider something like $\sum \rho_{aa} + \frac{1}{2} \sum_{|a-b|=1} \rho_{ab}$ or its obvious variants. These measures may also be justified easily by loss-function arguments.

9. PROPORTIONAL PREDICTION

Instead of basing a measure of association on optimal prediction one might consider measures based upon a prediction method which reconstructs the population, in a sense to be described. The use of such a measure was suggested to us by W. Allen Wallis. For simplicity, we restrict ourselves to the asymmetric situation of Section 5.1 where λ_b was constructed. Of course one could apply the same approach in other situations.

Our model of activity, as before, is the following: An individual is chosen at random from the population and we are asked to guess his B class either (1) given no information or (2) given his A class.

Optimal guessing will lead to a definite B class in case (1) and to a definite B class for each A class in case (2) (except that in the case of tied $\rho_{\cdot b}$'s or ρ_{ab}'s we have some choice). While such optimal guessing leads to the lowest average frequency of error, the resulting distribution of guessed classes will usually be very different from the original distribution in the population. For some purposes this might be undesirable and one is led to the following model of activity:

Case *1*. Guess B_1 with probability $\rho_{\cdot 1}$, B_2 with probability $\rho_{\cdot 2}$, $\cdots$, B_β with probability $\rho_{\cdot\beta}$.

Case *2*. Guess B_1 with probability $\rho_{a1}/\rho_{a\cdot}$ (the conditional probability of B_1 given A_a), B_2 with probability $\rho_{a2}/\rho_{a\cdot}$, $\cdots$, B_β with probability $\rho_{a\beta}/\rho_{a\cdot}$.

In each case the guessing is to proceed by throwing a β-sided die whose bth side appears with probability $\rho_{\cdot b}$ (case 1) or $\rho_{ab}/\rho_{a\cdot}$ (case 2). This may be accomplished using a table of "random numbers." If we make many such guesses independently it is plain that we shall approximately reconstruct the marginal distribution of the B_b's (case 1) and the joint distribution of the (A_a, B_b)'s (case 2).

The long-run proportion of *correct* predictions in case (1) will be $\sum_{b=1}^{\beta} \rho_{\cdot b}^2$, and in case (2) it will be $\sum_{a=1}^{\alpha} \sum_{b=1}^{\beta} \rho_{ab}^2/\rho_{a\cdot}$. Hence the relative *decrease* in the proportion of incorrect predictions as we go from case (1) to case (2) is

$$\tau_b = \frac{\sum_a \sum_b \rho_{ab}^2/\rho_{a\cdot} - \sum_b \rho_{\cdot b}^2}{1 - \sum_b \rho_{\cdot b}^2} \tag{31}$$

which can be readily expressed in the chi-square-like form

$$\tau_b = \frac{\sum_a \sum_b \frac{(\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2}{\rho_{a\cdot}}}{1 - \sum_b \rho_{\cdot b}^2}, \tag{32}$$

It is clear that τ_b takes values between 0 and 1; it is 0 if and only if there is independence, and 1 if and only if knowledge of A_a completely determines B_b. Finally τ_b is indeterminate if and only if both independence and determinism simultaneously hold, that is if all $\rho_{\cdot b}$'s but one are zero.

10. ASSOCIATION WITH A PARTICULAR CATEGORY

A group of modifications of many of the preceding measures arises from the observation that there may be little association between the A and B polytomies in general, but if an individual is in a particular A class it may be easy to predict his B class. Suppose, then, that we want the association between A_{a_0}, a specific A class, and the B polytomy. One need only condense all the A_a rows where $a \neq a_0$ into a single row, thus obtaining a $2 \times \beta$ table, and apply whatever measure of association is thought appropriate. The table will have this appearance.

	B_1	B_2	$\cdots$	B_β
A_{a_0}	$\rho_{a_0 1}$	$\rho_{a_0 2}$	$\cdots$	$\rho_{a_0 \beta}$
$A_a \ (a \neq a_0)$	$\rho_{\cdot 1} - \rho_{a_0 1}$	$\rho_{\cdot 2} - \rho_{a_0 2}$	$\cdots$	$\rho_{\cdot \beta} - \rho_{a_0 \beta}$

We are indebted to L. L. Thurstone for pointing out to us the importance of this modification.

11. PARTIAL ASSOCIATION

When there are more than two polytomies it is natural to think of partial association between two of them with the effect of the others averaged out in some sense. Two such measures of partial association will be suggested here for the asymmetrical case and three polytomies. The viewpoint will be that of optimal prediction. Analogous symmetrical measures may be readily obtained, and the restriction to three polytomies is purely for convenience of notation. The first two polytomies will be denoted as before; the third will consist of the classification $C_1, C_2, \cdots, C_\gamma$. The proportion of the population in A_a, B_b, and

C_c is ρ_{abc}, and dots will be used to denote marginal values in the conventional way. The proposed measures will be for partial association between the A and B polytomies 'averaged' over the C polytomy. (Do not confuse the integer γ used here with the index γ of Section 6.)

11.1. *Simple Average of* λ_b

For fixed C_c, we have a conditional $A \times B$ double polytomy with relative frequencies $\rho_{abc}/\rho_{\cdot\cdot c}$. Hence we can compute λ_b for each such table—call it $\lambda_b(c)$ to show dependence on c. Now it might seem natural to average these values with weights equal to the marginal relative frequencies of the C classifications. That is, we suggest

$$\lambda_b(A, B \mid C) = \sum_{c=1}^{\gamma} \rho_{\cdot\cdot c}\lambda_b(c). \tag{33}$$

11.2. *Measure Based Directly on Probabilities of Error*

It seems to us somewhat better, from the viewpoint of interpretation, to proceed as follows. For given C_c if we predict B classes optimally on the basis of no further information, the probability of error is $1-(\text{Max}_b\ \rho_{\cdot bc})/\rho_{\cdot\cdot c}$; whereas if we know the A class the probability of error is $1-(\sum_a \text{Max}_b\ \rho_{abc})/\rho_{\cdot\cdot c}$. Hence, if we are given individuals from the population at random and always told their C class, the probability of error in optimal guessing if we know nothing more is $1-\sum_c \text{Max}_b\ \rho_{\cdot bc}$; whereas if we also know the A class the probability is $1-\sum_c \sum_a \text{Max}_b\ \rho_{abc}$. Thus the relative decrease in probability of error is

$$\lambda_b'(A, B \mid C) = \frac{\sum_c \sum_a \text{Max}_b\ \rho_{abc} - \sum_c \text{Max}_b\ \rho_{\cdot bc}}{1 - \sum_c \text{Max}_b\ \rho_{\cdot bc}}, \tag{34}$$

which might often be a satisfactory measure of partial association.

12. MULTIPLE ASSOCIATION

When there are more than two polytomies one may well be interested in the multiple association between one of them and all the others. One simple way of handling this in the unordered case will be described here for three polytomies A, B, and C as defined in Section 11. We suppose that the multiple association between A and B-together-with-C is of interest. Simply form a two-way table whose rows represent the A polytomy and whose columns represent all combinations B_b, C_c and

then apply the appropriate two-polytomy measure. The table will have this appearance:

	B_1C_1	B_1C_2	$\cdots$	B_1C_γ	B_2C_1	$\cdots$	B_2C_γ	$\cdots$	$B_\beta C_\gamma$
A_1	ρ_{111}	ρ_{112}	$\cdots$	$\rho_{11\gamma}$	ρ_{121}	$\cdots$	$\rho_{12\gamma}$	$\cdots$	$\rho_{1\beta\gamma}$
A_2	ρ_{211}	ρ_{212}	$\cdots$	$\rho_{21\gamma}$	ρ_{221}	$\cdots$	$\rho_{22\gamma}$	$\cdots$	$\rho_{2\beta\gamma}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
A_α	$\rho_{\alpha 11}$	$\rho_{\alpha 12}$	$\cdots$	$\rho_{\alpha 1\gamma}$	$\rho_{\alpha 21}$	$\cdots$	$\rho_{\alpha 2\gamma}$	$\cdots$	$\rho_{\alpha\beta\gamma}$

Note that this procedure does not take the $B \times C$ association into account. There is a rough analogy here with the motivation for the standard multiple correlation coefficient of normal theory. The standard multiple correlation coefficient may be (and often is) motivated by defining it as the maximum correlation coefficient obtainable between a given variate and linear combinations of the other variates. That is, it is a measure of association between a given variate and the best estimate (in a certain sense) of that variate based upon all the other variates. It is true that the standard multiple correlation coefficient may be expressed as a function of the several ordinary bivariate correlation coefficients, but in a sense this is a consequence of the strong structural assumption of multivariate normality.

13. SAMPLING PROBLEMS

The discussion thus far has been in terms of *known* populations, whereas in practice one generally deals with a sample from an *unknown* population. One then asks, given a formal measure of association, how to estimate its value, how to test hypotheses about it, and so on.

Exact sampling theory for estimators from cross-classification tables is difficult to work with. However, the asymptotic theory is reasonably manageable, at least in some cases. We intend to discuss this in another paper, where we shall state some of the asymptotic distributions and say what we can of their value as approximations.

14. CONCLUDING REMARKS

The aim of this paper has been to argue that measures of association should not be taken blindly from the handiest statistics textbook, but rather should be carefully constructed in a manner appropriate to the problem at hand. To emphasize and illustrate this argument we have described a number of such measures which we feel might be useful in several situations. While we naturally take a friendly view towards these measures, we can hardly claim that they are more than examples.

This methodologically neutral position should not be carried to an extreme. It would be ridiculous to ask each empirical scientist in each separate study to forge afresh new statistical tools. The artist cannot paint many pictures if he must spend most of his time mixing pigments. Our belief is that each scientific area that has use for measures of association should, after appropriate argument and trial,[4] settle down on those measures most useful for its needs.

15. REFERENCES

[1] Cramér, Harald, *Mathematical Methods of Statistics*, Princeton, Princeton University Press (1946)

[2] Fisher, R. A., *Statistical Methods for Research Workers*, Tenth Edition, New York, Hafner Publishing Co. (1948)

[3] Greenwood, Major, Jr., and Yule, G. Udny, "The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general," *Proceedings of the Royal Society of Medicine*, 8 [part 2] (1915), 113–94.

[4] Guttman, Louis, "An outline of the statistical theory of prediction," Supplementary Study B-1 (pp. 253–318) in Horst, Paul and others (editors), *The Prediction of Personal Adjustment*, Bulletin 48, Social Science Research Council, New York (1941).

[5] Jahn, Julius A., "The measurement of ecological segregation: derivation of an index based on the criterion of reproducibility," *American Sociological Review*, 15 (1950), 101–4.

[6] Kendall, M. G., "Rank and product-moment correlation," *Biometrika*, 36 (1949), 177–93.

[7] Kendall, Maurice G., *The Advanced Theory of Statistics*, London, Charles Griffin and Co., Ltd. (1948).

[8] McCormick, Thomas C., "Toward causal analysis in the prediction of attributes," *American Sociological Review*, 17 (1952), 35–44.

[9] Pearson, Karl, and Heron, David, "On theories of association," *Biometrika*, 9 (1913), 159–315.

[10] Pompilj, G., "Osservazioni sull'omogamia: La trasformazione di Yule e il limite della trasformazione ricorrente di Gini," *Rendiconti di Matematica e*

[4] For examples of such argument and trial in the field of sociology see J. J. Williams [13], Jahn [5], and McCormick [8].

delle sue Applicazioni, Università di Roma, Istituto Nazionale di Alta Matematica, Ser. V, Vol. 9 (1950), 367–88.

[11] Stuart, A., "The estimation and comparison of strengths of association in contingency tables," *Biometrika*, 40 (1953), 105–10.

[12] Williams, E. J., "Use of scores for the analysis of information in contingency tables," *Biometrika*, 39 (1952), 274–89.

[13] Williams, Josephine J., "Another commentary on so-called segregation indices," *American Sociological Review*, 13 (1948), 298–303.

[14] Youden, W. J., "Index for rating diagnostic tests," *Cancer*, 3 (1950), 32–5.

[15] Yule, G. Udny, "On the methods of measuring association between two attributes," *Journal of the Royal Statistical Society*, 75 (1912), 579–642.

[16] Yule, G. Udny, and Kendall, M. G., *An Introduction to the Theory of Statistics*, London, Charles Griffin and Co., Ltd. (1950)

[17] Whelpton, P. K., and Kiser, Clyde V., *Social and Psychological Factors Affecting Fertility*, Milbank Memorial Fund, New York.

Volume 1 (1946), *The Household Survey in Indianapolis.*

Volume 2 (1950), *The Intensive Study; Purpose, Scope, Methods, and Partial Results.*

Volume 3 (1952), *Further Reports on Hypotheses in the Indianapolis Study.*

[18] Jaffe, A. J., review of [17], Volume 2, *Journal of the American Statistical Association*, 47 (1952), 348–9.

[19] Lewis-Faning, E., review of [17], Volume 3, *Journal of the American Statistical Association*, 49 (1954), 190–3.

A TEST OF GOODNESS OF FIT

T. W. Anderson and D. A. Darling*
Columbia University and University of Michigan

Some (large sample) significance points are tabulated for a distribution-free test of goodness of fit which was introduced earlier by the authors. The test, which uses the actual observations without grouping, is sensitive to discrepancies at the tails of the distribution rather than near the median. An illustration is given, using a numerical example used previously by Birnbaum in illustrating the Kolmogorov test.

1. THE PROCEDURE

The problem of statistical inference considered here is to test the hypothesis that a sample has been drawn from a population with a specified continuous cumulative distribution function $F(x)$. For example, the population may be specified by the hypothesis to be normal with mean 1 and variance $\frac{1}{6}$; the corresponding cumulative distribution function is

$$F(x) = \sqrt{\frac{3}{\pi}} \int_{-\infty}^{x} e^{-3(y-1)^2} dy. \tag{1}$$

In practice the procedure really tests the hypothesis that the sample has been drawn from a population with a completely specified density function, since the cumulative distribution function is simply the integral of the density.

The test procedure we have proposed earlier [1] is the following: Let $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ be the n observations in the sample in order, and let $u_i = F(x_i)$. Then compute

$$W_n^2 = -n - \frac{1}{n} \sum_{j=1}^{n} (2j-1)\left[\log u_j + \log(1 - u_{n-j+1})\right] \tag{2}$$

where the logarithms are the natural logarithms. If this number is too large, the hypothesis is to be rejected.

This procedure may be used if one wishes to reject the hypothesis whenever the true distribution differs materially from the hypothetical and especially when it differs in the tails.

Significance points for W_n^2 are not available for small sample sizes. The asymptotic significance points are given below:

* Work sponsored by Office of Scientific Research, U. S. Air Force, Contract AF18(600)-442, Project No. R-345-20-7.

The authors wish to acknowledge the assistance of Vernon Johns in the computations.

ASYMPTOTIC SIGNIFICANCE POINTS

Significance Level	Significance Point
.10	1.933
.05	2.492
.01	3.857

2. A NUMERICAL ILLUSTRATION

Birnbaum [2] has considered a sample of 40 observations and applied the Kolmogorov statistic to test the hypothesis that the population from which the data came was normal with mean 1 and standard deviation $1/\sqrt{6}$. By this test he found the data were consistent with the hypothesis. We have analyzed the same data using (2), obtaining $W_n^2 = 1.158$ which is well below the 10 per cent significance point, and we do not reject the hypothesis.

The computation sheet for this calculation had the following columns: x_j, $\sqrt{6}(x_j - 1)$, $u_j = F(x_j)$, $1 - u_{n-j+1}$, $\log u_j$, $\log(1 - u_{n-j+1})$ and $-[\log u_j + \log(1 - u_{n-j+1})]$. The operation $u_j = F(x_j)$ is simply finding the probability to the left of $\sqrt{6}(x_j - 1)$ according to the standard normal distribution.

Another test procedure uses the Cramer-von Mises ω^2 criterion given by

$$n\omega^2 = \frac{1}{12n} + \sum_{j=1}^{n} \left(u_j - \frac{2j-1}{2n} \right)^2. \tag{3}$$

The asymptotic distribution of this statistic is given in [1]. For Birnbaum's data we obtain $n\omega^2 = .1789$, which is also well below the 10 per cent asymptotic significance point of .3473

In these two examples we have used the asymptotic percentage points instead of the actual ones based on finite sample size. Empirical study suggests that the asymptotic value is reached very rapidly, and it appears safe to use the asymptotic value for a sample size as large as 40.

Application to the same data of the usual χ^2 criterion of K. Pearson, using 8 categories each with expected frequency 5, shows that $\chi^2 = 6.4$ which with 7 degrees of freedom is not significant at the 10 per cent level.

3. DERIVATION OF THE CRITERION

Several test procedures are based on comparing the specified cumulative distribution function $F(x)$ with its sample analogue, the empirical cumulative distribution function

$$F_n(x) = \frac{\text{no. of } x_i \leq x}{n}.$$

The present writers suggested [1] the use of the criterion

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi(F(x)) dF(x) \tag{4}$$

where $\psi(u)$ is some nonnegative weight function chosen by the experimenter to accentuate the values of $F_n(x) - F(x)$ where the test is desired to have sensitivity. The hypothesis is to be rejected if W_n^2 is sufficiently large. When $\psi(u) \equiv 1$ this criterion is n times the ω^2 criterion.

The criterion W_n^2 is an average of the squared discrepancy $[F_n(x) - F(x)]^2$, weighted by $\psi[F(x)]$ and the increase in $F(x)$ (and the normalization n). If one wishes the test to have good power against alternatives in which $H(x)$, the true distribution, and $F(x)$ disagree near the tails of $F(x)$, and to this end is willing to sacrifice power against alternatives in which $H(x)$ and $F(x)$ disagree near the median of $F(x)$, it seems that one ought to choose $\psi(u)$ to be large for u near 0 and 1, and small near $u = \frac{1}{2}$. Even if the alternative hypotheses are closely delineated, however, it appears difficult to find an "optimum" weight function $\psi(u)$. For a discussion of the general nature of power of distribution-free tests, see, for example, Birnbaum [3] and Lehmann [4].

For a given value of x, $F_n(x)$ is a binomial variable; it is distributed in the same way as the proportion of successes in n trials, where the probability of success is $H(x)$. Thus, $E[F_n(x)] = H(x)$ and

$$\begin{aligned} nE[F_n(x) - F(x)]^2 &= nE[F_n(x) - H(x)]^2 + n[F(x) - H(x)]^2 \\ &= H(x)[1 - H(x)] + n[F(x) - H(x)]^2. \end{aligned} \tag{5}$$

Under the null hypothesis ($H(x) = F(x)$), the variance is $F(x)[1 - F(x)]$. In a sense, we would equalize the sampling error over the entire range of x by weighting the deviation by the reciprocal of the standard deviation under the null hypothesis, that is, by using

$$\psi(u) = \frac{1}{u(1-u)} \tag{6}$$

as a weight function. This function has the effect of weighting the tails heavily since this function is large near $u = 0$ and $u = 1$. It is this weight function (6) which we treat in the present note.

Formula (2) is obtained by writing (4) as

$$
\begin{aligned}
\frac{1}{n} W_n^2 &= \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} dF(x) \\
&= \int_{-\infty}^{x_1} \frac{F^2(x)}{F(x)[1 - F(x)]} dF(x) + \int_{x_1}^{x_2} \frac{[F_n(x) - F(x)]^2}{F_n(x)[1 - F(x)]} dF(x) \\
&\qquad + \cdots + \int_{x_n}^{\infty} \frac{[1 - F(x)]^2}{F(x)[1 - F(x)]} dF(x),
\end{aligned}
$$

and letting $F(x) = u(dF(x) = du)$. Straightforward integration and collection of terms gives (2). The formula (2.5) in [1] cannot be used directly here, for that formula requires that $\int_0^1 \psi(u)du < \infty$, which is not true of (6).

4. COMPUTATION OF THE ASYMPTOTIC SIGNIFICANCE POINTS

It was proved in [1] that the limiting characteristic function of W_n^2 defined in either (2) or (4) is

$$
\phi(t) = \lim_{n\to\infty} E(e^{itW_n^2}) = \sqrt{\frac{-2\pi it}{\cos\left(\frac{\pi}{2}\sqrt{1+8it}\right)}} \tag{7}
$$

and that the inversion of this characteristic function gave for the limiting cumulative distribution of W_n^2 the expression

$$
\frac{\sqrt{2}}{z} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(j+\frac{1}{2})(4j+1)}{\cdots\, j!} e^{-[(4j+1)^2\pi^2]/(8z)} \cdot \int_0^{\infty} e^{z/[8(w^2+1)] - [(4j+1)^2\pi^2 w^2]/(8z)} \cdot dw. \tag{8}
$$

The terms of this series alternate in sign and the $(j+1)$st term is less than the jth term, $j \geqq 1$; thus the error involved in using only j terms of this series is less than the $(j+1)$st term for $j \geqq 0$. By using the fact that $e^{z/[8(w^2+1)]} \leqq e^{z/8}$, one can easily verify that to compute the probabilities correctly to four decimal places, one needs only the 0-th term for the first two significance points and the 0-th and 1-st terms for the third significance point. The laborious part of the computation is the evaluation of the integral. Let $[(4j+1)\pi/2\sqrt{z}]w = y$; then the integrand is $f(y)e^{-\frac{1}{2}y^2}$. The y-axis was divided into intervals according to the integral $e^{-\frac{1}{2}y^2}$ and numerical integration was performed.

The moments of the asymptotic distribution are fairly easy to obtain from formulas given in [1]. The first two are

$$\lim_{n\to\infty} E(W_n^2) = E(W_\infty^2) = \sum_{j=1}^{\infty} \frac{1}{j(j+1)} = 1,$$

$$\lim_{n\to\infty} \operatorname{Var}(W_n^2) = 2 \sum_{j=1}^{\infty} \frac{1}{j^2(j+1)^2} = \frac{2}{3}(\pi^2 - 9) \sim .57974.$$

The asymptotic significance points are computed to assure the probabilities (significance levels) to be correct to four decimal places.

REFERENCES

[1] Anderson, T. W., and Darling, D. A., "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Annals of Mathematical Statistics*, 23 (1952), 193–212.
[2] Birnbaum, Z. W., "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size," *Journal of the American Statistical Association*, 47 (1952), 425–41.
[3] Birnbaum, Z. W., "Distribution-free tests of fit for continuous distribution functions," *Annals of Mathematical Statistics*, 24 (1953), 1–8.
[4] Lehmann, E. L., "The power of rank tests," *Annals of Mathematical Statistics*, 24 (1953), 23–43.

UNIVARIATE TWO-POPULATION DISTRIBUTION-FREE DISCRIMINATION

DAVID S. STOLLER
*University of California at Los Angeles**

A distribution-free procedure for classifying a univariate random variable, z, into one of two populations on the basis of a sample of size N, in which m members are classified into one population and the remaining $(N-m)$ into the other, is given as follows: Let $t(z)=k(z)-h(z)$, where $k(z)$ is the number of observations from the first population which are less than z and $h(z)$ is similarly defined for the second population. If $z \leq \zeta^*$, where ζ^* is that value of z for which $t(z)$ is a maximum, classify z into the first population, otherwise into the second. The probability of correct classification, and its estimate, $[N-m+t(\zeta^*)]/N$, both converge in probability to the maximum attainable probability of correct classification.

1. INTRODUCTION

THE following example typifies a discrimination problem which is of interest. A group of N students take an aptitude test for a certain course, and receive scores $z_1, \cdots, z_N$. At the end of the course, all students are classified into two groups, "superior" and "inferior," on the basis of final grades or any other criteria. Another student takes the aptitude test and gets a score z_{N+1}. It is desired to classify this student as "superior" or "inferior." This will be done on the basis of selecting a discriminating score, ζ^*, based on the previous N scores, and classifying the student as "superior" if $z_{N+1} > \zeta^*$ and "average" if $z_{N+1} \leq \zeta^*$.

When there is complete *a priori* knowledge about the distribution functions and the relative frequency of occurrence of the two groups, Hoel and Peterson [6] have shown how to find an optimum discriminating point, ζ, by maximizing the probability of correctly classifying z_{N+1}. If the relative frequency of the two groups is not known, but there is otherwise complete knowledge, Anderson [1] and Welch [9] have shown how to find an optimum discriminating point by a minimax procedure.

If, further, there exists only partial *a priori* knowledge about the probability distributions, specifically, if the functional forms of the distributions are known but the parameters and relative frequencies are unknown, then, under certain restrictions, estimating the optimum discriminating point by replacing unknown parameters with their maximum likelihood estimates is an asymptotically optimum procedure [6].

* Now at The Rand Corporation.

A distribution-free procedure for the case where there exists no *a priori* knowledge of the parameters or form of the distribution functions has been investigated by Fix and Hodges [4], whereby z_{N+1} is classified into one group or another according to whether the sample values, z_i, "closest" to z_{N+1} are mostly in one or another group. In [4], consistency properties are proved about the probability of misclassification induced by this procedure. The small-sample behavior for some special cases is considered in a later paper [5].

The present paper proposes a distribution-free procedure for the univariate two-population case, together with an estimate of the probability of correct classification. It is shown that (1) the estimate of the probability of correct classification is a consistent estimate of the optimum probability of correct classification, (2) the probability of correct classification induced by this procedure converges in probability to the optimum probability of correct classification.

2. STATEMENT OF PROBLEM

Let Π be a composite univariate population, in which Π_1 and Π_2 are sub-populations with cumulative distribution functions $F_1(z)$ and $F_2(z)$. Let θ be the probability that z, a random member of Π, is a member of Π_1, i.e., z is a random variable defined by the cumulative distribution function, $\theta F_1(z)+(1-\theta)F_2(z)$.

A random sample, $z_1, \cdots, z_N$, is taken from Π, in which m of the z_i are identifiable as members of Π_1, and the remaining $N-m$ as members of Π_2. Another sample value, z_{N+1}, which is unidentifiable, is obtained at random. Without *a priori* knowledge of θ or of the functional forms or parameters of $F_1(z)$ and $F_2(z)$, it is desired to:

(*1*) Classify z_{N+1} as a member of Π_1 or Π_2.
(*2*) Estimate the probability that z_{N+1} has been correctly classified.

The functions $F_1(z)$ and $F_2(z)$, however, will be restricted to be such that (1) $F_1(z)$, $F_2(z)$ are absolutely continuous, so that the probability of tied sample values is zero), and (2) the optimum discrimination scheme consists of classifying z_{N+1} according to whether $z_{N+1} \leqq \zeta$ or $z_{N+1} > \zeta$, where ζ is a unique point. An optimum discrimination scheme is here defined as one that maximizes the probability of correct classification.

The probability of correct classification when an arbitrary point, z, is used to classify z_{N+1} by the above rule is given by:

$$Q(z) = \theta F_1(z) + (1 - \theta)[1 - F_2(z)].$$

By definition, $Q(z)$ achieves its maximum at $z = \zeta$.

3. A DISTRIBUTION-FREE DISCRIMINATION PROCEDURE

An estimate of the discriminating point, ζ, will be made by first considering a distribution-free estimate of the function, $Q(z)$. This estimate is formed by replacing θ by m/N, and $F_1(z)$ and $F_2(z)$ by the appropriate step functions, as follows.

Let $x_1, \cdots, x_m$ be the sample members from Π_1, *ordered by magnitude*, and $y_1, \cdots, y_{N-m}$ the similarly ordered members from Π_2. Define the step functions:

$$S_m^{(1)}(z) = k/m, \quad x_k \leqq z < x_{k+1}; \quad k = 0, \cdots, m,$$

$$S_{N-m}^{(2)}(z) = h/N - m, \quad y_h \leqq z < y_{h+1}; \quad h = 0, \cdots, N - m,$$

where

$$x_0 = y_0 = -\infty; \qquad x_{m+1} = y_{N-m+1} = +\infty.$$

Then a distribution-free estimate of $Q(z)$ is defined by

$$\begin{aligned}\widehat{Q}(z) &= (m/N)S_m^{(1)}(z) + [(N-m)/N][1 - S_{N-m}^{(2)}(z)] \\ &= \frac{1}{N}[(N-m) + (k-h)].\end{aligned}$$

(Note that $0 \leqq \widehat{Q}(z) \leqq 1$.) Take any value of z that maximizes $\widehat{Q}(z)$, say $z = \zeta^*$. Then ζ^* is an estimate of the optimum classification point, ζ, and $\widehat{Q}(\zeta^*)$ is an estimate of the probability of correct classification of the scheme: Classify z_{N+1} as a member of Π_1 if $z_{N+1} \leqq \zeta^*$, otherwise as a member of Π_2.

4. ILLUSTRATION

To illustrate the procedure, consider the following example. A class of 25 beginning algebra students took a test on elementary algebraic operations after two weeks of instruction. At the end of the course, the instructor classified the 25 students into two groups: "inferior" and "superior." Ranking the students by means of their two-week test scores, the following ordering resulted, where the scores of the 8 "superior" students are indicated by italics:

50, 51, 57½, 63, 64, 68, 72, 73, *74*, 74½, *75*, 75½, *76*,
76½, 77, 78, 81, 82, *84*, *85*, *86*, 87½, *89*, *91*, 92.

To this corresponds the following ordering of the x's and y's:

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, y_1, x_9, y_2, x_{10}, y_3,$$
$$x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, y_4, y_5, y_6, x_{16}, y_7, y_8, x_{17}.$$

Notice that $(k-h)$ may be computed very rapidly by the following procedure. At x_0, $(k-h)=0$. Proceeding along the ordered x's and y's, add one whenever an x is encountered, subtract one whenever a y is encountered. In the example, it can be seen that $(k-h)$ attains a maximum (of 12) in the interval from x_{16} to y_4, where $82 \leq z < 84$. Therefore, any point in this interval, say $\zeta^* = 83$, yields a discriminating procedure. An estimate of the probability of correct classification induced by $\zeta^* = 83$ is $\hat{Q}(83) = 20/25$, where the numerator is equal to $(N-m)$ plus the maximum value of $(k-h)$ and the denominator is N.

When applying the procedure described above, it may happen that two or more intervals exist in which $(k-h)$ is maximized. It is subsequently proved that any point which maximizes $(k-h)$ possesses asymptotically optimum characteristics. From a large sample point of view, therefore, when more than one maximizing interval is encountered, any point from any one of the maximizing intervals may be selected as ζ^*. In small samples, as a practical consideration, one should select an "average" value of ζ^*, say the average value of the midpoints of the maximizing intervals. The "average" value of ζ^*, as compared to any one value of ζ^*, will possess a more stable sampling behavior for small sample sizes, and will have the same asymptotically optimum characteristics discussed subsequently.

It may also happen that tied sample values will occur, due either to rounding off of observations or to *a priori* discreteness of the populations. In that category of ties which does not affect the calculation of ζ^*, the tied sample values may be ranked arbitrarily. When the calculation of ζ^* is affected by the occurrence of one or more sets of ties due to rounding off of observations, each critical set of ties in each population may be distributed uniformly over the round-off range of the observations, and then ordered accordingly, For example, in the sample:

$$1,\ 1,\ \mathit{1},\ 2,\ 3,\ 4,\ 4,\ \mathit{4},\ \mathit{4},\ \mathit{4},\ \mathit{5},\ \mathit{6},\ \mathit{7},$$

the tied sample values $(1, 1, \mathit{1})$, may be ranked arbitrarily since, by inspection, this does not affect the value of ζ^*. The set, $(4, 4, \mathit{4}, \mathit{4}, \mathit{4})$, which does affect the calculation of ζ^*, contains two observations from Π_1, which when distributed uniformly over the round-off range, $3\frac{1}{2}$ to $4\frac{1}{2}$, are assigned the values, $(3\frac{5}{6}, 4\frac{1}{6})$. Likewise, the set, $(\mathit{4}, \mathit{4}, \mathit{4})$, is assigned the values, $(3\frac{3}{4}, 4, 4\frac{1}{4})$.

When the populations are *a priori* discrete, and critical sets of tied values occur, $k(z)$ may be redefined as the number of observations from Π_1 less than or equal to z, and $h(z)$ as the number of observations from Π_2 less than or equal to z. Then $k(z) - h(z)$ will be uniquely defined even

when ties occur. In the example above, treating the sample values as *a priori* discrete, $k(1)-h(1)=1$, $k(4)-h(4)=2$, and $k(z)-h(z)$ is maximized in the interval, $3 \leqq z < 4$.

In Section 1, the criterion for determining an optimum classification scheme is based on maximizing the probability of correctly classifying one observation whose population membership is unknown. It should be noted that a classification scheme which is optimum in this sense for one unidentified observation is also optimum for a group of r independent, unidentified observations. For, if p represents the probability that one observation is correctly classified, then p^r is the probability that the entire group is correctly classified, and the latter probability is maximized when p is maximized.

5. DISTRIBUTION OF $\hat{Q}(z)$

For fixed m, h and k are independent binomial variates of means $mF_1(z)$ and $(N-m)F_2(z)$ respectively. Also, m is a binomial variate of mean $N\theta$, therefore it can be shown (by use of conditional expectation) that, for each z,

$$E[\hat{Q}(z)] = Q(z)$$

Thus, for each z, $\hat{Q}(z)$ is an unbiased estimate of $Q(z)$. It can also be shown that

$$\mathrm{Var}\,[\hat{Q}(z)] = c/N,$$

where c is a constant depending on θ, $F_1(z)$, $F_2(z)$, but not on N.

Since

$$\lim_{N\to\infty} \sigma_{\hat{Q}(z)} = 0,$$

by a Tchebychev inequality (see Cramér [2], Theorem 20.4) it is seen that, for each z, $\hat{Q}(z)$ is also a consistent estimate of $Q(z)$.

6. SOME PROPERTIES OF THE CLASSIFICATION PROBABILITY ESTIMATE, $\hat{Q}(\zeta^*)$

It is readily seen that $\hat{Q}(\zeta^*)$, the estimate of the classification probability induced by the point estimate, ζ^*, is non-negatively biased, since $\hat{Q}(\zeta^*) \geqq \hat{Q}(\zeta)$, and thus

$$E[\hat{Q}(\zeta^*) - \hat{Q}(\zeta)] \geq 0,$$

from which,

$$E[\hat{Q}(\zeta^*)] \geqq Q(\zeta).$$

An example (suggested by a referee) of the magnitude of the bias for small samples is given by the special case: $\theta=\frac{1}{2}; F_1(z)=F_2(z)$. Here $Q(\zeta)=\frac{1}{2}$, but, by definition, $Q(\zeta^*)\geqq\frac{1}{2}$. Further, if $m=N/2$,

$$\widehat{Q}(\zeta^*) = \tfrac{1}{2} + \tfrac{1}{2} \max_z (k/m - h/m)$$

and

$$\begin{aligned}&\Pr\{\widehat{Q}(\zeta^*) - Q(\zeta) > \beta\} = \\ &\Pr\{|\widehat{Q}(\zeta^*) - Q(\zeta)| > \beta\} = \\ &\Pr\{\max_z |k/m - h/m| > 2\beta\},\end{aligned}$$

the latter probability having been tabulated by Massey [8]. For two equal samples of size 10 from the same population, $\Pr\{\widehat{Q}(\zeta^*)\geqq .7\}$ equals about .16, and $\Pr\{\widehat{Q}(\zeta^*)\geqq .75\}$ equals about .05.

However, $\widehat{Q}(\zeta^*)$ is a consistent estimate of $Q(\zeta)$, for given ϵ, $\eta>0$, by sec. 5, for sufficiently large N,

$$\Pr\{\widehat{Q}(\zeta) < Q(\zeta) - \epsilon\} < \eta$$

and since $\widehat{Q}(\zeta^*)\geqq\widehat{Q}(\zeta)$,

$$\Pr\{\widehat{Q}(\zeta^*) < Q(\zeta) - \epsilon\} < \eta.$$

Now,

$$\begin{aligned}\widehat{Q}(\zeta^*) - Q(\zeta) = &\left\{\frac{N-m}{N} - (1-\theta)\right\} + \max_z \left\{\frac{k-h}{N}\right\} \\ &- \max_z \{\theta F_1(z) - (1-\theta)F_2(z)\} \leqq \left\{\theta - \frac{m}{N}\right\} \\ &+ \max_z \left\{\left[\frac{k}{N} - \theta F_1(z)\right] - \left[\frac{h}{N} - (1-\theta)F_2(z)\right]\right\} \\ \leqq &\left|\theta - \frac{m}{N}\right| + \max_z \left\{\frac{k}{N} - \theta F_1(z)\right\} \\ &+ \max_z \left\{\frac{h}{N} - (1-\theta)F_2(z)\right\}.\end{aligned}$$

Since m/N is a consistent estimate of θ, for $N>N'(\epsilon, \eta)$, the inequality,

$$\Pr\{|\theta - m/N| > \epsilon\} < \eta,$$

is satisfied. Now,

$$\max_z \left\{\frac{k}{N} - \theta F_1(z)\right\}$$

$$\leq \max_z \left\{\frac{k}{N} - \theta \frac{k}{m}\right\} + \max_z \left\{\theta \frac{k}{m} - \theta F_1(z)\right\}$$

$$\leq \left|\frac{m}{N} - \theta\right| + \max_z \left|\frac{k}{m} - F_1(z)\right|.$$

Consider the expression,

$$\max_z |k/m - F_1(z)|,$$

with m temporarily held fixed. Now

$$\sqrt{m} \max_z |k/m - F_1(z)|$$

is well known (Kolmogorov [7], Feller [3]) to possess an asymptotic probability distribution function with finite mean and variance. Therefore, for any fixed $m > M(\epsilon, \eta/2)$,

$$\Pr\left\{\max_z \left|\frac{k}{m} - F_1(z)\right| > \epsilon\right\} < \eta/2.$$

Now for $N > N''(M)$, $Pr\{m < M\} < \eta/2$, therefore for $N > N''(M)$,

$$\Pr\left\{\max_z \left|\frac{k}{m} - F_1(z)\right| > \epsilon\right\} < \eta.$$

A similar $N'''(\epsilon, \eta)$ exists for

$$\max_z \left|\frac{h}{m} - F_2(z)\right|;$$

thus for $N > \max(N', N'', N''')$,

$$\Pr\{\hat{Q}(\zeta^*) > Q(\zeta) + 5\epsilon\} < \eta.$$

7. AN ASYMPTOTICALLY OPTIMUM PROPERTY OF THE DISCRIMINATING POINT ESTIMATE, ζ^*

It can also be shown that $Q(\zeta^*)$, the actual (but unknown) classification probability resulting from the use of the discriminating point estimate, ζ^*, converges in probability to $Q(\zeta)$, the optimum classification probability, i.e. for sufficiently large sample sizes, $Q(\zeta^*)$ is arbitrarily close to $Q(\zeta)$ with probability artitrarily close to one. For,

$$|Q(\zeta^*) - Q(\zeta)| \leq |Q(\zeta^*) - \hat{Q}(\zeta^*)| + |\hat{Q}(\zeta^*) - Q(\zeta)|.$$

By sec. 6, for $N > N(\epsilon, \eta)$,

$$\Pr\{|\hat{Q}(\zeta^*) - Q(\zeta)| > \epsilon\} < \eta.$$

Now,

$$|Q(\zeta^*) - \hat{Q}(\zeta^*)| \leqq \left|\frac{m}{N} - \theta\right| + \left|\theta F_1(\zeta^*) - \frac{k}{N}\right| + \left|(1-\theta)F_2(\zeta^*) - \frac{h}{N}\right|.$$

Examining the second term,

$$\begin{aligned}\left|\theta F_1(\zeta^*) - \frac{k}{N}\right| &\leqq \left|\theta F_1(\zeta^*) - \theta\frac{k}{m}\right| + \left|\theta\frac{k}{m} - \frac{k}{N}\right| \\ &\leqq \theta \max_z \left|F_1(z) - \frac{k}{m}\right| + k\left|\frac{\theta}{m} - \frac{1}{N}\right| \\ &\leqq \max_z \left|F_1(z) - \frac{k}{m}\right| + \left|\theta - \frac{m}{N}\right|.\end{aligned}$$

Consequently, in a manner similar to the discussion of sec. 6, it can be shown that there exists an $N(\epsilon, \eta)$ such that for $N > N(\epsilon, \eta)$,

$$\Pr\{|Q(\zeta^*) - Q(\zeta)| > \epsilon\} < \eta.$$

REFERENCES

[1] Anderson, T. W., "Classification by multivariate analysis," *Psychometrika*, 16 (1951), 31–49.

[2] Cramér, H., *Mathematical Methods of Statistics*, Princeton, N. J., Princeton University Press (1946), p. 253.

[3] Feller, W., "On the Kolmogorov-Smirnov limit theorems for empirical distributions," *Annals of Mathematical Statistics*, 19 (1948), 177–89.

[4] Fix, Evelyn, and Hodges, J. L., *Discriminatory Analysis; Non-parametric Discrimination; Consistency Properties*, USAF School of Aviation Medicine, Randolph Field, Texas, Report Number 4, February, 1951.

[5] Fix, Evelyn, and Hodges, J. L., *Discriminatory Analysis; Non-parametric Discrimination; Small Sample Performance*, USAF School of Aviation Medicine, Randolph Field, Texas, Report Number 11, August, 1951.

[6] Hoel, P. G., and Peterson, R. P., "A solution to the problem of optimum classification," *Annals of Mathematical Statistics*, 20 (1949), 433–8.

[7] Kolmogorov, A. N., "Sulla determinazione empirica delle leggi di probabilita," *Giornale dell'Institute Italiano degli Attuari*, 4 (1933), 1–11.

[8] Massey, Frank J., "The distribution of the maximum deviation between two sample cumulative step functions," *Annals of Mathematical Statistics*, 22 (1951), 125–8.

[9] Welch, B. L., "Note on discriminant functions," *Biometrika*, 31 (1939), 218–20.

USE OF NORMAL PROBABILITY PAPER

HERMAN CHERNOFF AND GERALD J. LIEBERMAN
*Stanford University**

Normal probability paper is so designed that the cumulative distribution function of a normally distributed chance variable appears as a straight line. It is a common practice to plot the observations of a sample on this paper to obtain a graphical check for normality or to obtain a graphical estimate of the mean and variance of the population. Textbooks, however, are not very specific about methods for plotting, for, although the ordered observations are plotted along the abscissa, some uncertainties about the corresponding ordinates are left unresolved. The purpose of this paper is to indicate, with a special example, that any graphical technique should depend to a large extent on the purpose for which the graph is drawn. In particular, it presents tables covering sample sizes up to 10, for selecting the ordinates on normal probability paper so as to obtain "optimum" graphical estimates of the mean ξ and the standard deviation σ of a normal distribution. The somewhat more complicated problem of selecting the ordinates to obtain an "optimum" test for normality is not discussed.

1. UNBIASED ESTIMATES OF ξ AND σ

BY MEANS of a non-linear transformation of the vertical scale on the graph of the cumulative-normal-distribution curve, it is possible to transform this curve to a straight line. Graph paper possessing this property is known as normal probability paper. The abscissa scale corresponds to the values of a normally distributed chance variable, whereas the ordinate scale represents a number, p, between 0 and 1. Neither 0 nor 1 appears on the ordinate scale.

If a sample of n independent observations is to be plotted on normal probability paper, it is natural to arrange them in ascending order, i.e., $u_1 \leqq u_2 \leqq, \cdots \leqq u_n$, and to plot a point corresponding to each observation. One such plot is $(u_1, 1/n)$, $(u_2, 2/n)$, $\cdots$, $(u_n, n/n)$. However, it is evident that the last point does not appear on the graph. Furthermore, the symmetry of the normal distribution suggests that u_1 and u_n be treated in a "symmetric" fashion. Two alternative plots are

$$\left(u_1, \frac{1}{n+1}\right), \left(u_2, \frac{2}{n+1}\right), \cdots, \left(u_n, \frac{n}{n+1}\right)$$

* Work sponsored by the Office of Naval Research.

and

$$\left(u_1, \frac{1}{2n}\right) \left(u_2, \frac{3}{2n}\right), \left(u_3, \frac{5}{2n}\right), \cdots, \left(u_n, \frac{2n-1}{2n}\right).$$

Since there is no obvious rationale for preferring one of these plots to the other, there arises the problem of selecting an "optimum" method of plotting.

Let us consider the problem of graphically estimating the mean ξ and standard deviation σ of a normal population on the basis of a sample. Once the points $(u_1, p_1), (u_2, p_2), \cdots, (u_n, p_n)$ are plotted, a method which suggests itself is to fit a straight line visually to the points and to take the abscisa where the line intersects $p=.5$ as an estimate of the mean, and the distance between the abscissas where the line intersects $p=.8413$ and $p=.5$ as an estimate of the standard deviation. *Let us assume that the visually fitted line is a very good approximation to the line that would be obtained by minimizing the sum of squares of the horizontal deviations from the line.*[1] The problem then is to find what values of $p_1, p_2, \cdots, p_n$ yield good estimates, $\hat{\xi}$ and $\hat{\sigma}$, of the mean ξ and standard deviation σ of the normal population sampled. Since p is not represented on a linear scale, we shall transform to $v=v(p)$ which is related to p by

$$p = \int_{-\infty}^{v} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{1}$$

In terms of the ordinate v, the fitted straight line may be represented by

$$u = \hat{\xi} + \hat{\sigma} v \tag{2}$$

where $\hat{\xi}$ and $\hat{\sigma}$ are the estimates of ξ and σ. If $v_i=v(p_i)$, $i=1, 2, \cdots, n$, these estimates are

$$\hat{\xi} = \bar{u} - \hat{\sigma}\bar{v} \tag{3}$$

and

$$\begin{aligned} \hat{\sigma} &= \sum_{i=1}^{n} (u_i - \bar{u})(v_i - \bar{v}) \Big/ \sum_{i=1}^{n} (v_i - \bar{v})^2 \\ &= \sum_{i=1}^{n} u_i(v_i - \bar{v}) \Big/ \sum_{i=1}^{n} (v_i - \bar{v})^2. \end{aligned} \tag{4}$$

[1] The use of horizontal deviations is suggested by the fact that the p_i are not chance variables and the u_i are.

If we require that $\hat{\xi}$ be unbiased and $\hat{\sigma} \geqq 0$, we must have $\bar{v}=0$. In that case $\hat{\xi}=\bar{u}$ which is the mean of the sample. In fact, it is the optimum estimate of ξ from many points of view.

The estimate of σ may be represented by a linear function of the ordered observations $u_1, u_2, \cdots, u_n$, i.e.,

$$\hat{\sigma} = \sum_{i=1}^{n} c_i u_i \tag{5}$$

where

$$c_i = \frac{v_i - \bar{v}}{\sum_{i=1}^{n} (v_i - \bar{v})^2}, \qquad i = 1, 2, \cdots, n. \tag{6}$$

Let us define an optimum choice of the v_i to be one which minimizes the variance of $\hat{\sigma}$ subject to the condition that $\hat{\sigma}$ is an unbiased estimate of σ. We first note that if each v_i is increased by a constant, the c_i are unaffected. Hence we may select the v_i so that $\bar{v}=0$ without interfering with the optimum choice of v_i for estimating σ. Next we note that the class of all unbiased linear estimates of the ordered observations are available by suitably varying the v_i. The reason for this is that the only condition imposed by equation (3) on the c_i is $\sum_{i=1}^{n} c_i = 0$. This condition is implied by unbiasedness since

$$E\left(\sum_{i=1}^{n} c_i u_i\right) = k\sigma + \left(\sum_{i=1}^{n} c_i\right)\xi, \tag{7}$$

where k depends on $c_1, c_2, \cdots, c_n$. It follows that the problem of finding the optimum v_i is equivalent to that of finding the minimum variance unbiased estimate of σ which is linear in the ordered observations.

This problem is one which was treated by Godwin [1]. He presents a table of coefficients to be used with the ordered observations to obtain the minimum variance unbiased estimate of σ. His results were transformed for use in this paper. The "optimum" values of p_i can be found in Table I.

2 BIASED ESTIMATES OF σ

A peculiarity of the problem of estimating σ is that by introducing a slight bias we are able to obtain a better estimate in the sense of minimizing the mean square deviation from σ.

Suppose t is an estimate of θ where $E(t)=\theta$. Consider the statistic at

for which a is defined such that $E\{(at-\theta)^2\}$ is a minimum. In other words, we minimize the expression

$$E\{[(at - a\theta) + (a - 1)\theta]^2\} = a^2\sigma_t^2 + (a - 1)^2\theta^2$$

with respect to a, where σ_t^2 is the variance of t. This minimum occurs at

$$a = \frac{\theta^2}{\theta^2 + \sigma_t^2} \cdot$$

It then follows that

$$E\{(at - \theta)^2\} = \frac{1}{\dfrac{1}{\theta^2} + \dfrac{1}{\sigma_t^2}} \cdot$$

In the problem under consideration, t is a linear function of the ordered observations whose coefficients add up to zero. Hence σ_t^2 (for the given sample size) is a multiple of σ^2, say, $k_n\sigma^2$. Furthermore, $\theta = \sigma$, and therefore, the optimum a is given by

$$a = \frac{1}{1 + k_n} \cdot$$

The corresponding mean square deviation about σ is given by

$$E\{(at - \sigma)^2\} = \frac{\sigma^2}{1 + \dfrac{1}{k_n}} \cdot$$

Consider two unbiased estimates, linear in the ordered observations, which have variances $k_n\sigma^2$ and $k_n^*\sigma^2$ where $k_n < k_n^*$. The estimate with variance $k_n\sigma^2$ yields the better biased estimate (of the above type) since

$$\frac{1}{1 + \dfrac{1}{k_n}} < \frac{1}{1 + \dfrac{1}{k_n^*}} \cdot$$

Hence multiplying the c_i contained in the solution of the problem in the preceding section by an appropriate factor, which is equivalent to dividing each v_i^* by a, leads to the estimate which is optimum in the following sense. Among all estimates which are linear in the ordered observations and whose bias is independent of ξ, it has minimum mean square deviation from σ.

3. TABLES

In Table I the p_i values corresponding to the following are presented:

1. the best *linear* (in the ordered observations) unbiased estimate of σ,
2. the best *linear* (in the ordered observations) biased estimate of σ,
3. $p_i = \dfrac{i}{n+1}$,

and

4. $p_i = \dfrac{i - \frac{1}{2}}{n}$,

for values of $n = 1(1)10$.

TABLE I

COMPARISON OF THE ORDINATES (p_i) USED ON NORMAL PROBABILITY PAPER FOR ESTIMATING THE MEAN AND STANDARD DEVIATION

n		p_1	p_2	p_3	p_4	p_5
2	[1]	.28632				
	[2]	.18775				
	[3]	.33333				
	[4]	.25000				
3	[1]	.19870	.50000			
	[2]	.14020	.50000			
	[3]	.25000	.50000			
	[4]	.16667	.50000			
4	[1]	.14913	.40034			
	[2]	.10982	.38288			
	[3]	.20000	.40000			
	[4]	.12500	.37500			
5	[1]	.11775	.33333	.50000		
	[2]	.089398	.31272	.50000		
	[3]	.16667	.33333	.50000		
	[4]	.10000	.30000	.50000		
6	[1]	.096373	.28489	.42964		
	[2]	.074907	.26485	42229		
	[3]	.14286	.28571	.42857		
	[4]	.083333	.25000	.41667		

n		p_1	p_2	p_3	p_4	p_5
7	[1]	.080456	.24824	.37660	.50000	
	[2]	.063732	.22986	.36624	.50000	
	[3]	.12500	.25000	.37500	.50000	
	[4]	.071429	.21429	.35714	.50000	
8	[1]	.069624	.21959	.33500	.44544	
	[2]	.056027	.20289	.32348	.44139	
	[3]	.11111	.22222	.33333	.44444	
	[4]	.062500	.18750	.31250	.43750	
9	[1]	.060607	.19659	.30146	.40162	.50000
	[2]	.049425	.18158	.28978	.39537	.50000
	[3]	.10000	.20000	.30000	.40000	.50000
	[4]	.055556	.16667	.27778	.38889	.50000
10	[1]	.053568	.17773	.27386	.36559	.45537
	[2]	.044192	.16422	.26245	.35818	.45281
	[3]	.090909	.18182	.27273	.36364	.45455
	[4]	.050000	.15000	.25000	.35000	.45000

NOTE: When $i > n/2$, use $p_i = 1 - p_{n-i+1}$.

[1] These values of p_i correspond to the ordinates that yield the minimum variance unbiased estimate of σ which is linear in the ordered observations.

[2] These values of p_i correspond to the ordinates that yield the biased estimate of σ which has minimum mean square deviation from σ and which is linear in the ordered observations.

[3] These values of p_i correspond to $i/(n+1)$.

[4] These values of p_i correspond to $(i-\frac{1}{2})/n$.

Table II presents the mean square deviations from σ of the estimates which are linear in the ordered observations, the variance of the minimum variance non-linear unbiased estimate of σ, and the mean square deviation from σ of the non-linear biased estimate having minimum mean square deviation. The minimum variance non-linear unbiased estimate of σ[1] is

$$\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}\sqrt{\sum (x_i - \bar{x})^2}$$

and the corresponding biased estimate is

$$\frac{1}{1+k_n}\left[\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}\sqrt{\sum (x_i - \bar{x}^2)}\right].$$

This latter estimate is optimum in the sense that among all estimates having an expected value which is a multiple of σ and a variance proportional to σ^2, it has minimum mean square deviation from σ. This estimate

TABLE II

COMPARISON OF THE MEAN SQUARE DEVIATIONS FROM σ OF VARIOUS ESTIMATES OF σ

n	[1]	[2]	[3]	[4]	[5]	[6]
2	.57080	.57084	.36338	.36340	1.07533	.42611
3	.27324	.27549	.21460	.21599	.49856	.22649
4	.17810	.18006	.15117	.15259	.31559	.15558
5	.13177	.13332	.11643	.11764	.22751	.11872
6	.10447	.10571	.09459	.09560	.17630	.09605
7	.08650	.08714	.07961	.08015	.14306	.08067
8	.07379	.07469	.06872	.06950	.11987	.06954
9	.06432	.06501	.06044	.06105	.10283	.06111
10	.05701	.05759	.05393	.05445	.08981	.05449

[1] Variance of the minimum variance non-linear unbiased estimate of σ.
[2] Variance of the minimum variance unbiased estimate which is linear in the ordered observations.
[3] Mean square deviation from σ of the non-linear biased estimate which has minimum mean square deviation.
[4] Mean square deviation from σ of the biased estimate which is linear in the ordered observations and which has minimum mean square deviation.
[5] Mean square deviation from σ of the biased estimate based upon the ordinates $i/(n+1)$.
[6] Mean square deviation from σ of the biased estimate based upon the ordinates $(i-\frac{1}{2})/n$.

is referred to above and subsequently as the non-linear biased estimate having minimum mean square deviation from σ.

In Table III these mean square deviations are transformed to efficiencies. For the unbiased estimates the ratio of the variances are computed; for the biased estimates the ratio of the mean square deviations from σ to the minimum are computed.

It is evident from Tables II and III that the optimum choice of the p_i depends upon whether an unbiased estimate is necessary or whether a biased estimate can be tolerated. In either case, the graphical estimates compare very favorably with the optimum estimates of the standard deviation. For $n \leqq 10$ the efficiency of the optimum unbiased graphical estimate relative to the optimum unbiased estimate is greater than 98 per cent, as is also the efficiency of the optimum biased graphical estimate relative to the optimum biased estimate.

TABLE III-A

EFFICIENCY OF THE OPTIMUM UNBIASED ESTIMATE OF σ (RATIO OF VARIANCES)

n	[1]
2	100.00
3	99.19
4	98.92
5	98.84
6	98.83
7	98.86
8	98.90
9	98.94
10	98.99

TABLE III-B

COMPARISON OF EFFICIENCY OF VARIOUS BIASED ESTIMATES OF σ (RATIO OF MEAN SQUARE DEVIATIONS FROM σ)

n	[2]	[3]	[4]
2	99.99	33.79	85.28
3	99.36	43.04	94.75
4	99.07	47.90	97.17
5	98.97	51.18	98.07
6	98.94	53.65	98.48
7	99.33	55.65	98.68
8	99.88	57.33	98.82
9	99.00	57.78	98.90
10	99.04	60.05	98.97

[1] This entry is the ratio of the variance of the minimum variance non-linear unbiased estimate to the variance of the minimum variance unbiased estimate which is linear in the ordered observations, i.e., columns [1]/[2] in Table II.

[2] This entry is the ratio of the mean square deviation from σ of the non-linear biased estimate having minimum mean square deviation to the mean square deviation from σ of the minimum mean square deviation biased estimate which is linear in the ordered observations, i.e., columns [3]/[4] in Table II.

[3] This entry is the ratio of the mean square deviation from σ of the non-linear biased estimate having minimum mean square deviation to the mean square deviation from σ of the biased estimate based upon the ordinates $i/(n+1)$, i.e., columns [3]/[5] in Table II.

[4] This entry is the ratio of the mean square deviation from σ of the non-linear biased estimate having minimum mean square deviation to the mean square deviation from σ of the biased estimate based upon the ordinates $(i-\frac{1}{2})/n$, i.e., columns [3]/[6] in Table II.

4. REFERENCES

[1] Godwin, H. J., "On the estimation of dispersion by linear systematic statistics," *Biometrika*, 36 (1949), 92–100.

[2] Godwin, H. J., "Some low moments of order statistics," *Annals of Mathematical Statistics* 20 (1949), 279–85.

ANALYSIS OF SIMPLE LATTICE DESIGNS WITH UNEQUAL SETS OF REPLICATIONS*

PAUL MEIER
The Johns Hopkins University

INTRODUCTION

THE lattice, or pseudo-factorial, designs first introduced by Yates [11] and various generalizations of them [7, 8] have proved to be quite useful, particularly in agricultural applications. These designs are suited to experimental situations in which there are a large number of varieties, treatments, or what have you, to be compared under conditions which require a relatively small block size.

Among the wide variety of available incomplete block designs the lattices are distinguished by the fact that they combine a relatively simple type of analysis with a fair degree of flexibility in the choice of the number of varieties to be tested, the number of replications to be used, and so forth.

In most of these designs the basic construction consists of p replication patterns. The instructions for analysis, such as given in [2], allow for any number of repetitions of any subset of this basic collection of replication patterns. Thus if we use p' patterns from the basic collection and repeat this subset r times we have a design with a total of rp' replications. For example, if we wish to compare 25 varieties in a 5×5 lattice design we have available a basic set of 6 patterns. An experiment with 6 replications may be designed using all six patterns once, using each of three patterns twice, or using just two patterns each repeated three times. (See Cochran and Cox [2, p. 281].) However, if we wish to use 5 replications our choice would be limited to the quintuple lattice, and the case of 7 replications is not covered at all. The possibilities are even further restricted in the case of a 6×6 lattice for which the basic set includes only 3 patterns.

Even when the above prescription can be followed it may not be the most desirable procedure. The more patterns from the basic set used, the more tedious becomes the analysis, so that in situations where the complexity and cost of calculations weigh heavily the designs using more patterns, although generally more balanced, may lose in favor. Conversely, it may be possible to achieve a more nearly balanced de-

* This paper is a revision of parts III and IV of a doctoral thesis submitted to Princeton University. Preparation of this paper was assisted by a contract with the Office of Naval Research. Paper No. 293 from the Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University.

sign by dropping the requirement that each pattern be repeated the same number of times.

Finally, although it may not be a frequent occurrence, there are unfortunate occasions on which for one reason or another essentially all of one or more replications is lost, thus destroying the symmetry of the original arrangement.

For these reasons it may be of interest to note the extent to which the ordinary analysis for lattice designs must be modified for an experiment in which the p' patterns from the basic set are repeated with *unequal* frequencies.

In this paper we investigate the simplest case, the simple square lattice, which uses only two replication patterns. Following the usual convention replications following one pattern are called X-replications and those following the other pattern are called Y-replications. We will refer to a design using n X-replications and m Y-replications as an (n, m) design.

In addition to the question of relative ease of analysis, the efficiency of unequal sets designs relative to alternatives is of interest and this is considered in section I–G. The criterion used to measure efficiency is the reciprocal of the average variance of all possible comparisons between varieties, which is a kind of average "information," or invariance, of comparisons. In experiments of identical construction having different numbers of replications this quantity will be proportional to the number of replications. Thus, to compare the efficiencies of two designs which use different numbers of replications we use the average invariance of comparisons divided by the number of replications. This gives us an absolute scale on which to compare any two designs involving the same number of varieties.

It is not claimed that this is an ideal criterion, but it is felt to be satisfactory for the purpose at hand, namely, to compare an unequal sets lattice design with an equal sets design. We will also compare the (2,1) design with the triple lattice and the (3, 2) design with the quintuple lattice. Since a 5×5 lattice is generally considered to be the smallest to which the recovery of inter-block information should be applied, we give numerical calculations for it. The disadvantage of the unequal sets lattice will be less in larger designs.

It will be seen that the analysis for the unequal sets simple lattice designs differs to only a slight extent from the usual analysis given for the equal sets designs. In view of this simplicity such designs may be considered legitimate competitors with the more nearly balanced alternatives, provided the relative efficiency is not too low. The maxi-

mum losses of efficiency for the (2, 1) and (3, 2) designs relative to the equal sets designs and the triple and quintuple lattices, respectively, are given below for the 5×5 lattice.

MAXIMUM LOSS IN EFFICIENCY FOR A 5×5 SIMPLE LATTICE WITH UNEQUAL SETS OF REPLICATIONS RELATIVE TO OTHER DESIGNS

Alternative Design	(2, 1) Design	(3, 2) Design
Simple Lattice with Equal Sets	6%	2%
Triple Lattice	12%	—
Quintuple Lattice	—	11%

These maximum losses are realized when block variability is large or the intra-block analysis is used.

The paper is divided into two parts. The first part deals with the theory of the analysis and the model behind it. The second part consists of a numerical example.

I. DESIGN AND ANALYSIS

A. Field Design and Mathematical Model

The simple lattice designs permit the comparison of k^2 varieties in blocks of size k. Thus a complete replication consists of k blocks, each containing k varieties. The construction of replication patterns begins by arranging the varieties in a square array. In the first, or X, pattern those varieties appearing in any one *row* go into the same block. In the second, or Y, pattern those varieties appearing in any one *column* go into the same block. In agricultural work the blocks thus composed would be assigned at random to the blocks laid out in a replication, and the varieties within a block would be assigned at random to the plots within a block. In other types of work the analogous method of assignment would be followed. An excellent description of this procedure is given by Cochran and Cox [2].

The mathematical model may be described as follows. Consider the varieties in the square array from which the design is composed. Denote by v_{ij} the "true" mean of the variety in the ith row and jth column (less the over-all mean of the varieties). Denote by x_{ijr} the value observed for the plot in the rth replication containing the variety v_{ij}. Then

$$x_{ijr} = \mu + A_r + v_{ij} + \alpha_{ir} + \beta_{jr} + \epsilon_{ijr}$$

where

μ = Grand mean
A_r = Replication effect
v_{ij} = Variety effect

$$\alpha_{ir} = \begin{cases} i\text{th block effect if the } r\text{th replication is in the } X \text{ set} \\ 0 \text{ if the } r\text{th replication is in the } Y \text{ set} \end{cases}$$

$$\beta_{jr} = \begin{cases} j\text{th block effect if the } r\text{th replication is in the } Y \text{ set} \\ 0 \text{ if the } r\text{th replication is in the } X \text{ set} \end{cases}$$

ϵ_{ijr} = Residual effect.

In the analysis that follows we will assume that the ϵ_{ijr} may be considered to be normally and independently distributed about zero with variance σ_ϵ^2. For the intra-block analysis we need no other assumptions since block effects are completely eliminated from the varietal comparisons. A direct application of Cochran's theorem [4] to the analysis of variance shows that the residual mean square is a proper estimate of error, and by a little additional calculation the analysis can also be made to provide an exact test of significance for the null-hypothesis that all v_{ij} are equal.

For the inter-block estimation of varietal effects we must make some assumptions about the block effects, α_{ir} and β_{jr}. Besides the assumption of randomness, insured by the method of assigning blocks within replications, it is necessary to assume that block variability is the same within each replication. To avoid expository clumsiness it is convenient to assume, when dealing with the inter-block analysis, that the block effects also are normally and independently distributed with variance σ_B^2. (The calculation of average mean squares requires only that the effects be uncorrelated, but this is not sufficient to justify the application of the t-distribution to varietal comparisons. Of course, strict normality need not be required. It has been shown that for many purposes deviations from normality do not seriously affect the analysis [5].) The quantities σ_R^2 and σ_v^2 which appear in the expressions for the average mean squares represent the variation due to replication and varietal effects. Assumptions about the nature of the distribution of these effects or their random allocation have no bearing on either type of analysis. Except in discussing the recovery of inter-block information, block effects will be considered fixed rather than random effects.

As are most of the familiar experimental designs, the unequal sets simple lattice is a partially balanced incomplete block (p.b.i.b.) design (see discussion in [2]). The application to particular designs of the general method for analyzing a p.b.i.b. design is nicely demonstrated in a paper by Nair [10]. However, the inequality in number of X- and Y-

replications in our case gives us a design with three rather than the usual two associate classes, and it would appear from this viewpoint that the unequal sets case is essentially more complex than the equal sets designs. However, the analysis follows directly by application of Cochran's theorem without making any appeal to general results from the theory of experimental designs. In this form the close similarity of the analysis to that for equal sets is obvious.

The problem of choosing a suitable notation for exposition is by no means simple. Our object is to provide a notation which suggests the operations involved without being unduly cumbersome. For the quantities which are sums or averages over all possible values of an index we use the fairly common convention that a dot (·) replacing an index means the average over all possible values of that index, whereas a plus (+) replacing an index means the sum over all possible values of that index. For example:

$$x_{i\cdot\cdot} = \frac{1}{kR} \sum_j \sum_r x_{ijr}; \qquad x_{ij+} = \sum_r x_{ijr}.$$

The only difficulty with this notation is that a mean taken over X replications, for example, takes the clumsy form

$$\frac{1}{R_1} \sum_{r=1}^{R_1} x_{ijr}$$

where we have, say, R_1-X-replications and R_2-Y-replications. To avoid writing such expressions we use the notation $\bar{x}_{ij}$ for the above, and $\bar{\bar{x}}_{ij}$ for a mean over Y-replications. It follows that $Rx_{ij.} = R_1\bar{x}_{ij} + R_2\bar{\bar{x}}_{ij}$ where $R = R_1 + R_2$ is the total number of replications.

B. Analysis of Variance

The analysis[1] follows from the identity

$$\sum_{ijr} (x_{ijr} - x_{\dots})^2 \qquad \textit{Total}$$

$$= k^2 \sum_r (x_{\cdot\cdot r} - x_{\dots})^2 \qquad \text{Replications}$$

$$+ R \sum_{ij} (x_{ij\cdot} - x_{\dots})^2 \qquad \text{Varieties ignoring blocks}$$

$$+ k\left\{ \sum_{r=1}^{R_1} \sum_i \left[(x_{i\cdot r} - x_{\cdot\cdot r}) - (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})\right]^2 \right. \qquad \text{Blocks A}$$

[1] Expressions suitable for computation are given with the numerical example in part II.

$$+ \sum_{r=R_1+1}^{R} \sum_{j} [(x_{\cdot jr} - x_{\cdot\cdot r}) - (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})]^2 \Big\}$$

Blocks B (eliminating varieties)

$$+ \frac{kR_1R_2}{R} \Big\{ \sum_{i} [(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}) - (\bar{\bar{x}}_{i\cdot} - \bar{\bar{x}}_{\cdot\cdot})]^2$$

$$+ \sum_{j} [(\bar{\bar{x}}_{\cdot j} - \bar{\bar{x}}_{\cdot\cdot}) - (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})]^2 \Big\}$$

Residual

$$+ \Big\{ \sum_{r=1}^{R_1} \sum_{ij} (x_{ijr} - x_{i\cdot r} - x_{ij\cdot} + x_{i\cdot\cdot} + x_{\cdot j\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\cdot\cdot} - x_{\cdot\cdot\cdot})^2$$

$$+ \sum_{r=R_1+1}^{R} \sum_{ij} (x_{ijr} - x_{\cdot jr} - x_{ij\cdot} + x_{i\cdot\cdot} + x_{\cdot j\cdot} - \bar{\bar{x}}_{i\cdot} + \bar{\bar{x}}_{\cdot\cdot} - x_{\cdot\cdot\cdot})^2 \Big\}.$$

It can be verified directly that the sums of cross products such as

$$\sum_{ijr} (x_{\cdot\cdot r} - x_{\cdot\cdot\cdot})(x_{ij\cdot} - x_{\cdot\cdot\cdot})$$

are all zero. From this it follows that the rank of the quadratic form is equal to the sum of the ranks of the various components. It follows from Cochran's theorem [4] that the component sums of squares are distributed independently. With the exception of the residual terms the ranks of the quadratic forms can be seen directly, and the rank of the residual may be obtained by subtraction.

Also, excepting the residual, the average mean squares are easy to determine. For example

$$\text{Ave} \Big\{ k^2 \sum_{r} (x_{\cdot\cdot r} - x_{\cdot\cdot\cdot})^2 \Big\}$$

$$= \text{Ave} \Big\{ k^2 [\sum_{r} (A_r - A_\cdot) + (\alpha_{\cdot r} + \beta_{\cdot r} - \alpha_{\cdot\cdot} - \beta_{\cdot\cdot}) + (\epsilon_{\cdot\cdot r} - \epsilon_{\cdot\cdot\cdot})]^2 \Big\}.$$

As the various effects are assumed to be independent, the average values of the cross terms are zero and the expression becomes

$$\text{Ave} \Big\{ \sum_{r} (A_r - A_\cdot)^2 + \sum_{r} (\alpha_{\cdot r} + \beta_{\cdot r} - \alpha_{\cdot\cdot} - \beta_{\cdot\cdot})^2$$

$$+ \sum_{r} (\epsilon_{\cdot\cdot r} - \epsilon_{\cdot\cdot\cdot})^2 \Big\}$$

$$= (R - 1)\, \sigma_R^2 + \frac{(R-1)}{k} \sigma_B^2 + \frac{(R-1)}{k^2} \sigma_e^2$$

and the average mean square for replications is $k^2/(R-1)$ times this expression. In the same manner we find the other average means squares, as usual, obtaining the residual by subtraction. The result of these calculations may be tabulated as follows.

Source of variation	Average mean square	Degrees of freedom
Replications	$k^2\sigma_R^2+k\sigma_B^2+\sigma_e^2$	$R-1$
Varieties ignoring blocks	$R\sigma_v^2+\frac{k}{k+1}\sigma_B^2+\sigma_e^2$	k^2-1
Blocks A	$k\sigma_B^2+\sigma_e^2$	$(R-2)(k-1)$
Blocks B (eliminating varieties)	$\frac{1}{2}k\sigma_B^2+\sigma_e^2$	$2(k-1)$
Residual	σ_e^2	$(k-1)(Rk-k-1)$
Total		Rk^2-1

C. *Tests of Significance*

The analysis of variance table just presented does not suggest an exact test of significance for differences between varietal means. It may be, especially in an experiment with a large number of varieties, that the experimenter well knows that differences do exist and the significance test is really beside the point. Cochran [3] has described rather completely both approximate and exact tests of significance in the case of equal sets of replications. The unequal sets case may be treated in the same way.

An approximate test may be made by viewing the experiment as a randomized complete block design. That is, we regard the variation due to blocks as part of the experimental error. This is given effect by pooling the blocks and residual sums of squares and using the pooled mean square to test the significance of the varieties ignoring blocks mean square.

If we regard the block effects as randomly drawn from a normal distribution with mean zero and variance σ_B^2, the distribution of the test ratio under the null hypothesis is that of the ratio of two mixtures of chi-squares, each with the same average value. Thus the null distribution is not quite the same as the F distribution. However, the error arising from this source appears to be quite small.

An exact test can be found from an alternative analysis of variance. The foregoing analysis can be further subdivided as follows. The sum

of squares for varieties ignoring blocks may be written

$$R\sum_{ij}(x_{ij\cdot}-x_{\cdots})^2 = kR\left\{\sum_i (x_{i\cdot\cdot}-x_{\cdots})^2+\sum_j (x_{\cdot j\cdot}-x_{\cdots})^2\right\} + R\sum_{ij}(x_{ij\cdot}-x_{i\cdot\cdot}-x_{\cdot j\cdot}+x_{\cdots})^2$$

or SS (varieties ignoring blocks)
$=SS$ (varietal main effects ignoring blocks) $+SS$ (varietal interactions). (The terminology is borrowed from the pseudo-factorial description of the analysis [11].)

The average mean squares and degrees of freedom are as follows.

Source of variation	Average mean square	Degrees of freedom
Varieties ignoring blocks	$R\sigma_v^2+\frac{k}{k+1}\sigma_B^2+\sigma_e^2$	k^2-1
Varietal main effects ignoring blocks	$R\sigma_v^2+\frac{1}{2}k\sigma_B^2+\sigma_e^2$	$2(k-1)$
Varietal interactions	$R\sigma_v^2+\sigma_e^2$	$(k-1)^2$

Now the SS (varietal main effects ignoring blocks) can be replaced by an expression for SS (varietal main effects eliminating blocks).

SS (varietal main effects eliminating blocks)

$$= kR_1\sum_j(\bar{x}_{\cdot j}-\bar{\bar{x}}_{\cdot\cdot})^2+kR_2\sum_i(\bar{\bar{x}}_{\cdot i}-\bar{\bar{x}}_{\cdot\cdot})^2.$$

The average mean square is $R\sigma_v^2+\sigma_e^2$ again on $2(k-1)$ degrees of freedom. Unfortunately this sum of squares is not independent[2] of blocks B. To restore independence we must replace blocks B by

$$\text{blocks } B_u = SS\text{ (blocks } B \text{ ignoring varieties)}$$
$$SS\text{ (blocks } B_u) = kR_1\sum_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2+kR_2\sum_j(\bar{\bar{x}}_{\cdot j}-\bar{\bar{x}}_{\cdot\cdot})^2.$$

The average mean square is

$$\tfrac{1}{2}R\sigma_v^2+k\sigma_B^2+\sigma_e^2$$

[2] For this reason, the present writer prefers to avoid presenting a single table including both varieties eliminating blocks and blocks eliminating varieties as in [10]. The degrees of freedom in such a table add up correctly, but the sums of squares do not.

on $2(k-1)$ degrees of freedom. An exact test of significance for varietal effects is now provided by the pooled mean square for varieties eliminating blocks (on k^2-1 degrees of freedom) against the residual. We give below the average mean squares corresponding to this analysis of variance table.

Source of variation	Average mean square	Degrees of freedom
Replications	$k^2\sigma_R^2+k\sigma_B^2+\sigma_e^2$	$R-1$
Varietal main effects (eliminating blocks)	$\frac{1}{2}R\sigma_v^2+\sigma_e$	$2(k-1)$
Varietal interactions	$R\sigma_v^2+\sigma_e^2$	$(k-1)^2$
Blocks A	$k\sigma_B^2+\sigma_e^2$	$(R-2)(k-1)$
Blocks B_u	$\frac{1}{2}R\sigma_v^2+k\sigma_B^2+\sigma_e^2$	$2(k-1)$
Residual	σ_e^2	$(k-1)(Rk-k-1)$

D. *Estimation of Varietal Means*

In the following we will use the notation v_{ij}' for the intra-block estimate of the true varietal mean, v_{ij}, and v_{ij}'' for the inter-block estimate. v_{ij}' is obtained by averaging the appropriate values, one from each replication, and adding two corrections to eliminate block effects, one for the X-replications and one for the Y-replications.

$$v_{ij}' = x_{ij\cdot} + C_{Xi} + C_{Yj}$$

where

$$C_{Xi} = \frac{R_1}{R}\left[(\bar{\bar{x}}_{i\cdot} - \bar{\bar{x}}_{\cdot\cdot}) - (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})\right]$$

and

$$C_{Yj} = \frac{R_2}{R}\left[(\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot}) - (\bar{\bar{x}}_{\cdot j} - \bar{\bar{x}}_{\cdot\cdot})\right].$$

An easy calculation shows that block effects are completely eliminated from this estimate. However, the elimination of block error necessarily introduces additional residual error. If a reasonable estimate of the ratio of block variance to residual variance is available, a partial correction will give an estimate with smaller variance. Thus, we consider an estimate of the form

$$v_{ij}'' = x_{ij\cdot} + \lambda C_{Xi} + \nu C_{Yj},$$

where λ and ν are chosen to minimize the variance of v_{ij}''. These minimizing values and the method of estimating them are discussed in the next two sections.

E. Variances of Varietal Differences

The varietal differences will be of three kinds, each having a different variance:

1) both varieties in the same X block
2) both varieties in the same Y block
3) varieties not in the same X or Y block.

The variances of varietal differences for arbitrarily fixed values of λ and ν can be calculated directly. For example, for varieties in the same X block we have

$$V_1 = \text{Var}\,\{v_{11}'' - v_{12}''\} = \frac{2\sigma_e^2}{Rk}\,(k + \frac{R_2}{R_1}\,\nu^2) + \frac{2(1-\nu)^2 R_2 \sigma_B^2}{R^2}.$$

It is easily seen that these variances are all minimized by the same values of λ and ν, namely

$$\lambda^* = \frac{R_2 k \sigma_B^2}{R_2 k \sigma_B^2 + R\sigma_e^2} \quad \text{and} \quad \nu^* = \frac{R_1 k \sigma_B^2}{R_1 k \sigma_B^2 + R\sigma_e^2},$$

and for these values of λ and ν the three variances reduce to

1) $$V_1 = \frac{2\sigma_e^2}{Rk}\left(k + \frac{R_2}{R_1}\,\nu^*\right).$$

2) $$V_2 = \frac{2\sigma_e^2}{Rk}\left(k + \frac{R_1}{R_2}\,\lambda^*\right)$$

3) $$V_3 = \frac{2\sigma_e^2}{Rk}\left(k + \frac{R_1}{R_2}\,\lambda^* + \frac{R_2}{R_1}\,\nu^*\right).$$

Since λ^* and ν^* lie between zero and one, these variances will not differ by much if k is large and R_1 and R_2 are nearly equal. In this case, the average variance, weighted with the number of comparisons of each type, may suffice. This is found to be

4) $$V_4 = \frac{2\sigma_e^2}{R(k+1)}\left(k + \frac{R_1}{R_2}\,\lambda^* + \frac{R_2}{R_1}\,\nu^* + 1\right).$$

In the above calculations the quantities λ^* and ν^* are treated as constants, whereas in fact they must be estimated from the experimental data. When k and R are both small, the error involved may be large. A preliminary investigation which will be reported later indicates that when $k>6$, the error will definitely be negligible.

F. Estimation of λ^* and ν^*

If the ratio of σ_B^2 to σ_e^2 were known, λ^* and ν^* could be determined exactly. In the absence of such knowledge, the ratio can be estimated from the residual and blocks eliminating varieties sums of squares. The optimum method of using this information has not been determined, but the method originally given by Yates [12] seems to be satisfactory. The blocks A and blocks B sums of squares are pooled and equated to the resulting average value. Proceeding similarly with the residual we now have two equations in the two unknowns, σ_B^2 and σ_e^2. The solutions of these two equations are substituted into the formulas for λ^* and ν^*. If it should happen that the pooled block mean square is actually smaller than the residual mean square, we take zero as our estimate of λ^* and ν^*. The resulting estimates of λ^* and ν^* are

$$\hat{\lambda}^* = \frac{b - e}{b + \dfrac{R_1 - 1}{R_2} e} \text{ if } b \geqq e \quad \text{and} \quad \hat{\lambda}^* = 0 \text{ if } b < e,$$

$$\hat{\nu}^* = \frac{b - e}{b + \dfrac{R_2 - 1}{R_1} e} \text{ if } b \geqq e \quad \text{and} \quad \hat{\nu}^* = 0 \text{ if } b < e,$$

where b = pooled mean square for blocks eliminating varieties, and e = residual mean square. For the inter-block analysis the variances of varietal differences are estimated as follows.

If $b \geqq e$, we substitute the estimates $\hat{\lambda}^*$ and $\hat{\nu}^*$ in place of λ^* and ν^* in the formulas on p. 795, and use the residual mean square as an estimate of σ_e^2. If $b<e$, we replace λ^* and ν^* by zero and use the pooled mean square for blocks and residual to estimate σ_e^2.

For the intra-block analysis there is no need to estimate λ^* and ν^*. We take $\lambda = \nu = 1$ in the formulas on p. 795 and use the residual mean square to estimate σ_e^2.

The intra-block estimates of the variances of varietal comparisons have the usual distribution of a mean square with the residual degrees

of freedom. The distribution of the inter-block estimates is complex, but it will be reasonably well approximated by a mean square distribution with the same degrees of freedom.

G. Efficiency of the Unequal Sets Designs

The usual measure of efficiency for an experiment designed to compare several similar quantities is the reciprocal of the variance of the estimated difference between any two of them. When, as is the case in lattice designs, different comparisons do not all have the same variance, the usual practice is to use the reciprocal of the average variance of all possible comparisons. Since the variance of a comparison in a given experiment is proportional to the number of replications used, the above quantity divided by the number of replications is a measure of the intrinsic efficiency of the design. We need such an intrinsic criterion particularly in order to gauge the efficiency of the (2, 1) and (3, 2) designs relative to a lattice with equal sets of replications.

The efficiency of a given design relative to an alternative will be measured by the ratio of the above criteria for the two designs in experiments for which σ_B^2 and σ_ϵ^2 are the same for both designs. The calculation of variances is straightforward and results in the following.

AVERAGE VARIANCE OF VARIETAL COMPARISONS FOR VARIOUS DESIGNS

Design	Average variance of comparisons
Randomized Blocks	$\dfrac{2\sigma_\epsilon^2}{R(k+1)}\left[k+k\,\dfrac{\sigma_B^2}{\sigma_\epsilon^2}+1\right]$
Lattice with Equal Sets	$\dfrac{2\sigma_\epsilon^2}{R(k+1)}\left[k+\dfrac{k\sigma_B^2}{\frac{1}{2}k\sigma_B^2+\sigma_\epsilon^2}+1\right]$
Triple Lattice	$\dfrac{2\sigma_\epsilon^2}{R(k+1)}\left[k+\dfrac{k\sigma_B^2}{\frac{2}{3}k\sigma_B^2+\sigma_\epsilon^2}+1\right]$
Quintuple Lattice	$\dfrac{2\sigma_\epsilon^2}{R(k+1)}\left[k+\dfrac{k\sigma_B^2}{\frac{4}{5}k\sigma_B^2+\sigma_\epsilon^2}+1\right]$

The triple and quintuple lattices are included for comparison with the (2, 1) and 3, 2) designs. Comparing the average variance of comparisons in the unequal sets lattice with the variances listed we find the following relative efficiencies.

EFFICIENCY OF UNEQUAL SETS LATTICE RELATIVE TO ALTERNATIVE DESIGNS

Alternative design	Relative efficiency* of (R_1, R_2) simple lattice
Randomized Blocks	$\dfrac{k+\dfrac{k\sigma_B^2}{\sigma_e^2}+1}{k+\dfrac{R_1}{R_2}\lambda^*+\dfrac{R_2}{R_1}\nu^*+1}$
Lattice with Equal Sets	$\dfrac{k+\dfrac{k\sigma_B^2}{\frac{1}{2}k\sigma_B^2+\sigma_e^2}+1}{k+\dfrac{R_1}{R_2}\lambda^*+\dfrac{R_2}{R_1}\nu^*+1}$
Triple Lattice	$\dfrac{k+\dfrac{k\sigma_B^2}{\frac{2}{3}k\sigma_B^2+\sigma_e^2}+1}{k+\dfrac{R_1}{R_2}\lambda^*+\dfrac{R_2}{R_1}\nu^*+1}$
Quintuple Lattice	$\dfrac{k+\dfrac{k\sigma_B^2}{\frac{4}{5}k\sigma_B^2+\sigma_e^2}+1}{k+\dfrac{R_1}{R_2}\lambda^*+\dfrac{R_2}{R_1}\nu^*+1}$

* These relative efficiencies are calculated without allowance for the inaccuracy of weighting. This discriminates against the randomized block design which is, in fact, somewhat more efficient than the alternatives (rather than slightly less) when σ_B^2 is very small.

The relative efficiencies using the intra-block analysis may be obtained by replacing λ^* and ν^* by one and, in the case of the alternative lattices, taking the limit as σ_B^2/σ_e^2 becomes large. For the randomized block design there is no intra-block analysis and the efficiency remains a function of σ_B^2/σ_e^2.

In the comparisons with other lattice designs it will be noted that the relative efficiencies approach one when k is large, or when σ_B^2 approaches zero and the inter-block analysis is used. It can also be verified by straightforward algebra that the least favorable values for the unequal sets designs occur when σ_B^2 is large, or when the intra-block analysis is used. Since $k=5$ is frequently taken as the lower limit for reasonable accuracy in the recovery of inter-block information, e.g. [2], we will examine this case in detail.

We note first that our unequal sets design is the least efficient of the lattices compared. This is only to be expected since it is the furthest from balance. However, compared with the equal sets design it does not do at all badly. In the worst case, that of large σ_B^2/σ_e^2, or the intra-

block analysis, the efficiency relative to the equal sets design becomes

$$\frac{k+3}{k+\dfrac{R_1}{R_2}+\dfrac{R_2}{R_1}+1}.$$

Thus, for a 5×5 lattice the (2, 1) design is at worst 94 per cent efficient and the (3, 2) design 98 per cent efficient. These efficiencies will improve with larger k, as noted earlier. Hence, to a fair approximation, the efficiency of unequal sets designs relative to alternatives is about the same as that of the equal sets designs, provided the inequality in the number of X- and Y-replications is not excessive.

The advantage of the unequal sets designs lies in the simplicity of the calculations required, these being no more arduous than for the lattice with equal sets. In the triple and quintuple lattices, on the other hand, the additional labor involved in finding the adjusted means may be considerable. In circumstances where the cost of analysis is an appreciable portion of the cost of experimentation, the unequal sets designs may be regarded as competitors of the more balanced, but more complex, triple and quintuple lattice designs. The least favorable situation for the unequal sets designs relative to the triple and quintuple lattices is again found to be the intra-block analysis with k small. We see directly that using the intra-block analysis with $k=5$, the (2, 1) design is 88 per cent efficient relative to the triple lattice and the (3, 2) design is 89 per cent efficient relative to the quintuple lattice.

If it is known in advance that the block variability is not excessive, we may be assured of somewhat greater efficiency. The usual measure of block variability is the ratio of the variance between blocks to that within blocks which is

$$\gamma=\frac{k\sigma_B^2+\sigma_e^2}{\sigma_e^2}$$

(or w/w' in the notation of [2] and [12]). We see that the efficiencies can be expressed as functions of γ. In the above two cases the maximum loss in efficiency will be reduced by at least half if γ is no greater than five.

II. NUMERICAL EXAMPLE[3]

The material used in the example is drawn from two separate ex-

[3] The writer is indebted for the experimental data to Dr. P. H. Harvey, Agronomist, Bureau of Plant Industry, U. S. Dept. of Agronomy, North Carolina State College, Raleigh, N. C. The material was transmitted to the writer through the courtesy of Dr. R. J. Monroe, North Carolina State College, Raleigh, N. C.

periments at different locations. Both experiments used the same varieties and design, each with two X- and two Y-replications. To construct an example of a (3, 2) lattice, an X-replication from the second experiment was added to the first experiment, giving three X- and two Y-replications. Any interpretation of our numerical results must bear in mind the artificial nature of the "experiment" analyzed. The crop in question was corn and the variable was field weight of grain in pounds. Each plot consisted of 30 ($=2\times15$) single plant hills.

Table I presents the basic data with the totals for each block. In Table II we have the X-replications combined and the Y-replications combined, with row and column totals in each case. Table III gives the grand total for each variety with the row and column totals.

TABLE I

SINGLE PLOT YIELDS

	X-REPLICATIONS							*SS*
X_1								
	14.2	16.0	14.8	16.7	16.8	16.3	94.8	
	14.6	15.8	16.0	14.9	12.5	15.2	89.0	
	17.3	19.8	19.4	10.3	13.5	15.6	95.9	
	16.0	13.6	15.3	11.6	14.4	15.4	86.3	
	13.2	12.3	15.2	13.8	15.3	13.1	82.9	
	16.7	14.6	15.8	17.0	17.9	13.5	95.5	
							544.4	83711.8
X_2								
	13.2	15.6	13.9	14.5	12.2	15.1	84.5	
	16.5	18.3	16.1	11.4	13.0	16.9	92.2	
	13.8	19.0	15.9	13.5	18.0	14.4	94.6	
	15.6	13.7	14.0	12.5	14.1	14.0	83.9	
	14.0	13.0	13.0	15.8	14.5	11.1	81.4	
	16.2	14.8	15.3	18.0	16.6	12.5	93.4	
							530.0	79332.8
X_3								
	14.2	18.3	19.3	14.3	18.3	14.9	99.3	
	15.5	16.9	21.0	12.6	12.8	14.2	93.0	
	11.3	20.1	15.4	16.7	18.8	17.1	99.4	
	16.2	15.8	15.6	9.8	18.0	16.0	91.4	
	12.2	10.3	16.3	12.7	15.7	15.3	82.5	
	16.1	14.5	18.7	19.7	16.3	12.9	98.2	
							563.8	90884.2

Y-REPLICATIONS

Y_1								
	17.3	15.0	19.5	19.6	17.7	20.0	109.1	
	18.7	18.8	20.2	19.3	15.8	17.0	109.8	
	16.4	23.2	19.1	16.5	16.9	16.7	108.8	
	18.0	22.8	15.5	13.2	18.3	19.0	106.8	
	15.1	14.0	14.0	16.2	12.6	17.6	89.5	
	18.1	16.0	13.2	16.5	14.0	15.1	92.9	
							616.9	107923.5
Y_2								
	14.6	15.8	15.7	15.3	13.1	16.9	91.4	
	14.0	16.1	19.4	16.8	12.5	18.2	97.0	
	17.5	22.5	15.4	19.7	18.6	17.8	111.5	
	16.5	18.9	12.8	14.8	16.3	17.7	97.0	
	18.5	15.1	18.1	17.1	16.6	16.4	101.8	
	19.6	18.2	14.0	17.5	15.1	14.1	98.5	
							597.2	100730.4

TABLE II

COMBINATION OF REPLICATIONS

$X = X_1 + X_2 + X_3$						Totals	RR_2kC_{Xi}
41.6	49.9	48.0	45.5	47.3	46.3	278.6	55.7
46.6	51.0	53.1	38.9	38.3	46.3	274.2	100.8
42.4	58.9	50.7	40.5	50.3	47.1	289.9	10.9
47.8	43.1	44.9	33.9	46.5	45.4	261.6	84.3
39.4	35.6	44.5	42.3	45.5	39.5	246.8	68.9
49.0	43.9	49.8	54.7	50.8	38.9	287.1	45.3
266.8	282.4	291.0	255.8	278.7	263.5	1638.2	365.9
$Y = Y_1 + Y_2$							RR_1kC_{Yj}
31.9	30.8	35.2	34.9	30.8	36.9	200.5	−67.9
32.7	34.9	39.6	36.1	28.3	35.2	206.8	−55.6
33.9	45.7	34.5	36.2	35.5	34.5	220.3	−78.9
34.5	41.7	28.3	28.0	34.6	36.7	203.8	−99.8
33.6	29.1	32.1	33.3	29.2	34.0	191.3	−16.5
37.7	34.2	27.2	34.0	29.1	29.2	191.4	−47.2
204.3	216.4	196.9	202.5	187.5	206.5	1214.1	−365.9

A. The Analysis of Variance

The analysis proceeds in the usual way as follows. The correction term is given by

$$C = \frac{(x_{+++})^2}{Rk^2} = \frac{(2852.3)^2}{180} = 45197.86.$$

The total sum of squares is obtained by summing the squares of each plot yield and subtracting the correction term.

$$\textit{Total SS} = \sum_{ijr} (x_{ijr})^2 - C = (14.2)^2 + (16.0)^2 + \cdots + (14.1)^2 - C$$

$$= 46258.27 - 45197.86 = 1060.41.$$

TABLE III

VARIETY TOTAL YIELDS

73.5	82.6	81.9	80.0	80.9	84.0	482.9
77.4	85.9	98.8	80.6	67.4	80.5	490.6
77.6	98.5	85.2	68.8	82.4	74.3	486.8
82.7	79.2	81.1	61.9	79.8	79.4	464.1
70.2	63.9	80.0	76.9	74.7	68.6	434.3
85.9	79.1	84.3	91.4	84.8	68.1	493.6
467.3	489.2	511.3	459.6	470.0	454.9	2852.3

Similarly,

$$\textit{Replications SS} = \frac{1}{k^2} \sum_{r} (x_{++r})^2 - C$$

$$= \frac{1}{36} [(544.4)^2 + (530.0)^2 + (563.8)^2 + (616.9)^2 + (597.2)^2] - C$$

$$= 45343.20 - 45197.86 = 145.34,$$

$$\textit{Varieties ignoring blocks SS} = \frac{1}{R} \sum_{ij} (x_{ij+})^2 - C$$

$$= \frac{1}{5} [(73.5)^2 + \cdots + (68.1)^2] - C$$

$$= 45667.87 - 45197.86 = 470.01.$$

The blocks A sum of squares is computed in two parts, one from the X-replications (blocks A′) and one from the Y-replications (blocks A″).

$$\begin{aligned}
\textit{Blocks } A' \, SS &= \frac{1}{k}\sum_i \sum_{r=1}^{R_1} (x_{i+r})^2 - \frac{1}{k^2}\sum_{r=1}^{R_1} (x_{++r})^2 \\
&\quad - \frac{1}{R_1 k}\sum_i \left(\sum_{r=1}^{R_1} x_{i+r}\right)^2 + \frac{1}{R_1 k^2}\left(\sum_{r=1}^{R_1} x_{++r}\right)^2 \\
&= \frac{1}{6}\,[(94.8)^2 + \cdots + (98.2)^2] \\
&\quad - \frac{1}{36}\,[(544.4)^2 + (530.0)^2 + (563.8)^2] \\
&\quad - \frac{1}{18}\,[(278.6)^2 + \cdots + (287.1)^2] \\
&\quad + \frac{1}{108}\,[1638.2]^2 \\
&= 13.77
\end{aligned}$$

X-REPLICATION BLOCK TOTALS

				Total
	94.8	84.5	99.3	278.6
	89.0	92.2	93.0	274.2
	95.9	94.6	99.4	289.9
	86.3	83.9	91.4	261.6
	82.9	81.4	82.5	246.8
	95.5	93.4	98.2	287.1
Total	544.4	530.0	563.8	1638.2

$$\begin{aligned}
\textit{Blocks } A'' \, SS &= \frac{1}{k}\sum_j \sum_{r=R_1+1}^{R} (x_{+jr})^2 - \frac{1}{k^2}\sum_{r=R_1+1}^{R} (x_{++r})^2 \\
&\quad - \frac{1}{R_2 k}\sum_j \left(\sum_{r=R_1+1}^{R} x_{+jr}\right)^2 + \frac{1}{R_2 k^2}\left(\sum_{r=R_1+1}^{R} x_{++r}\right)^2 \\
&= \frac{1}{6}\,[(109.1)^2 + \cdots + (98.5)^2] \\
&\quad - \frac{1}{36}\,[(616.9)^2 + (597.2)^2]
\end{aligned}$$

$$-\frac{1}{12}[(200.5)^2+\cdots+(191.4)^2]$$

$$+\frac{1}{72}[1214.1]^2$$

$$=58.20$$

Y-REPLICATION BLOCK TOTALS

			Total
	109.1	91.4	200.5
	109.8	97.0	206.8
	108.8	111.5	220.3
	106.8	97.0	203.8
	89.5	101.8	191.3
	92.9	98.5	191.4
Total	616.9	597.2	1214.1

In the case of blocks A″ we have only two replications which permits a simplification of the computation.

$$\text{Blocks } A'' \text{ } SS=\frac{1}{R_2k}\sum_j(x_{+j4}-x_{+j5})^2-\frac{1}{R_2k^2}(x_{++4}-x_{++5})^2$$

$$=\frac{1}{2\times 6}[(17.7)^2+(12.8)^2+\cdots+(5.6)^2]$$

$$-\frac{1}{2\times 36}(19.7)^2$$

$$=58.20$$

Y-REPLICATION BLOCK TOTALS

			Diff.
	109.1	91.4	17.7
	109.8	97.0	12.8
	108.8	111.5	− 2.7
	106.8	97.0	9.8
	89.5	101.8	−12.3
	92.9	98.5	− 5.6
Total	616.9	597.2	19.7

The blocks B sum of squares is also computed in two parts. It will be noted that the quantities to be squared are proportional to the unadjusted correction terms for blocks, C_{X_i} and C_{Y_j}. In this computation the terms in $\bar{x}_{..} - \bar{\bar{x}}_{..}$ may be ignored as the calculation corrects for means automatically. Thus in place of C_{X_i} and C_{Y_j} it is more convenient to use

$$C_{X_i}' = \frac{R_1}{R} (\bar{\bar{x}}_{i\cdot} - \bar{x}_{i\cdot}),$$

$$C_{Y_j}' = \frac{R_2}{R} (\bar{x}_{\cdot j} - \bar{\bar{x}}_{\cdot j}).$$

Thus

$$\begin{aligned} RR_2kC_{X_i}' &= R_1(R_2\bar{\bar{x}}_{i+}) - R_2(R_1\bar{x}_{i+}) \\ &= R_1(i\text{th column total for group } Y) \\ &\quad - R_2(i\text{th row total for group } X) \end{aligned}$$

and similarly,

$$\begin{aligned} RR_1kC_{Y_j}' &= R_2(j\text{th row total for group } X) \\ &\quad - R_1(j\text{th column total for group } Y). \end{aligned}$$

For example:

$$RR_2kC_{X_1}' = 3(204.3) - 2(278.6) = 55.7.$$

In this manner, we obtain:

$RR_2kC_{X_i}'$	$RR_1kC_{Y_j}'$
55.7	−67.9
100.8	−55.6
10.9	−78.9
84.3	−99.8
68.9	−16.5
45.3	−47.2
365.9	−365.9

As a computational check, it may be noted that the sum of the $RR_2kC_{X_i}'$ is just RR_1R_2 (sum of all X plots − sum of all Y plots). The sum of the $RR_1kC_{Y_j}'$ is the negative of this.

We may now write the sum of squares for blocks B.

$$\begin{aligned}
\textit{Blocks } B' \; SS &= \frac{1}{RR_1R_2k} \sum_i (RR_2kC_{X_i}')^2 - \frac{1}{RR_1R_2k^2} (RR_2kC_{X+}')^2 \\
&= \frac{1}{5 \times 3 \times 2 \times 6} [(55.7)^2 + (100.8)^2 + \cdots + (45.3)^2] \\
&\quad - \frac{1}{5 \times 3 \times 2 \times 6^2} (365.9)^2 \\
&= 27.63.
\end{aligned}$$

$$\begin{aligned}
\textit{Blocks } B'' \; SS &= \frac{1}{RR_1R_2k} \sum_j (RR_1kC_{Y_j}')^2 - \frac{1}{RR_1R_2k^2} (RR_1kC_{Y+})^2 \\
&= \frac{1}{5 \times 3 \times 2 \times 6} [(67.9)^2 + (55.6)^2 + \cdots + (47.2)^2] \\
&\quad - \frac{1}{5 \times 3 \times 2 \times 6^2} (365.9)^2 \\
&= 22.63.
\end{aligned}$$

The residual sum of squares may now be obtained by difference:

$$\begin{aligned}
\textit{Residual } SS &= 1060.41 - (145.34 + 470.01 + 13.77 + 58.20 + 27.63 + 22.63) \\
&= 322.83.
\end{aligned}$$

The above calculations give the analysis of variance table as follows.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Replications	4	145.34	36.335
Varieties ignoring blocks	35	470.01	13.429
Blocks A	15	71.97	4.798
Blocks B	10	50.26	5.026
Residual	115	322.83	2.807
Total	179	1060.41	

B. *Tests of Significance*

A simple approximate F-test for varietal effects is made by combining the blocks and residual sums of squares, giving an error term of

$$71.97 + 50.26 + 322.83 = 445.06 \text{ on } 140 \text{ d.f.}$$

or a mean square of 3.179. This is to be compared with varieties ignoring blocks, giving

$$F = \frac{13.429}{3.179} = 4.224 \text{ on 35 and 140 d.f.}$$

which is highly significant. The test is not exact because the error term is a mixture of mean squares. The error in significance level caused by using this approximateion has been investigated by Cochran [1]. In a case such as this the error will be quite small.

If an exact F-test is required, we must perform a little extra computation. We require the sum of squares for varieties eliminating blocks, which is most easily found from the identity

$$\left\{\begin{array}{l} SS \text{ (varieties eliminating blocks)} \\ +SS \text{ (blocks ignoring varieties)} \end{array}\right\} = \left\{\begin{array}{l} SS \text{ (varieties ignoring blocks)} \\ +SS \text{ (blocks eliminating varieties)} \end{array}\right\}.$$

Now SS (blocks ignoring varieties)

$$= \left[\frac{1}{k}\sum_i \sum_{r=1}^{R_1} (x_{i+r})^2 - \frac{1}{k^2}\sum_{r=1}^{R_1} (x_{++r})^2\right]$$

$$+ \left[\frac{1}{k}\sum_j \sum_{r=R_1+1}^{R} (x_{+jr})^2 - \frac{1}{k^2}\sum_{r=R_1+1}^{R} (x_{++r})^2\right]$$

$$= \frac{1}{6}[(94.8)^2 + \cdots + (98.2)^2] - \frac{1}{36}[(544.4)^2 + (530.0)^2 + (563.8)^2]$$

$$+ \frac{1}{6}[(109.1)^2 + \cdots + (98.5)^2] - \frac{1}{36}[(616.9)^2 + (597.2)^2]$$

$$= 195.19.$$

Hence SS (varieties eliminating blocks)

$$= 470.01 + 122.23 - 195.19 = 397.05 \text{ on 35 d.f.}$$

and the corresponding mean square is 11.344. This is to be tested against the residual mean square, giving

$$F = \frac{11.344}{2.807} = 4.012 \text{ on 35 and 115 d.f.,}$$

which is also highly significant.

C. Estimation of Varietal Means

If the intra-block analysis is to be used, the estimation of varietal means is now trivial. We need merely adjust the raw varietal means by adding the quantities C_{Xi} and C_{Yj}, i.e.

$$v_{ij}' = x_{ij\cdot} + C_{Xi} + C_{Yj}\,.$$

The use of C_{Xi}' and C_{Yj}' in place of C_{Xi} and C_{Yj} introduces a constant bias which does not affect the comparisons and is generally quite small. The bias is readily removed by adding the quantity $-1/k(C_{X+}' + C_{Y+}')$ which in our case is equal to -0.34.

To calculate the inter-block estimates of the varietal means, we must first estimate the coefficients λ^* and ν^*. The formulas for these estimates appear on p. 796 and reduce in this example to

$$\hat{\lambda}^* = \frac{b - e}{b + e}$$

$$\hat{\nu}^* = \frac{b - e}{b + \frac{1}{3}e}$$

where b = pooled mean square for blocks eliminating varieties and e = residual mean square. In this example

$$b = \frac{71.97 + 50.26}{25} = 4.889$$

and $e = 2.807$ so that

$$\hat{\lambda}^* = 0.2705 \quad \text{and} \quad \hat{\nu}^* = 0.3574.$$

Multiples of the unadjusted correction terms have already been calculated, namely, RR_2kC_{Xi}' and RR_1kC_{Yj}'. The adjusted correction terms are found by multiplying these quantities by

$$\frac{\hat{\lambda}^*}{RR_2k} = \frac{0.2705}{60} = 0.004503 \quad \text{and} \quad \frac{\hat{\nu}^*}{RR_1k} = \frac{0.3574}{90} = 0.003971,$$

respectively. (Replace $\hat{\lambda}^*$ and $\hat{\nu}^*$ by one for the intra-block estimates). These corrections are appended to the table of varietal means (Table IV) and the appropriate terms added to each raw varietal mean yielding the corrected means (Table V).

The additional adjustment due to using C_{Xi}' and C_{Yj}' in place of C_{Xi} and C_{Yj} is

$$-\frac{1}{k}(\hat{\lambda}^*C_{X+}' + \hat{\nu}^*C_{Y+}') = -0.03,$$

which may be added to each varietal mean, if desired.

It may be of interest to point out that the need for these corrections is peculiar to the unequal sets designs. The corrections are identically zero in an equal sets design.

TABLE IV

VARIETY MEANS (UNADJUSTED)

							$\hat{\lambda}C_{X_i}'$
	14.70	16.52	16.38	16.00	16.18	16.80	0.251
	15.48	17.15	19.76	16.12	13.48	16.10	0.454
	15.52	19.70	17.02	13.76	16.48	14.86	0.049
	16.54	15.84	16.22	12.38	15.96	15.88	0.380
	14.04	12.78	16.00	15.38	14.94	13.72	0.311
	17.18	15.82	16.86	18.28	16.96	13.62	0.204
$\hat{\nu}C_{Y_j}'$	−0.270	−0.221	−0.313	−0.396	−0.066	−0.187	

TABLE V

VARIETY MEANS (ADJUSTED)

14.681	16.550	16.318	15.855	16.365	16.864
15.664	17.383	19.901	16.178	13.868	16.367
15.299	19.528	16.756	13.413	16.463	14.722
16.650	15.999	16.287	12.364	16.274	16.073
14.081	12.870	15.998	15.296	15.185	13.844
17.114	15.803	16.751	18.088	17.098	13.637

D. *Variances and Standard Errors of Varietal Differences*

The variances of varietal differences depend on whether or not the varieties being compared appear together in some block. We give first the variances and standard errors for the intra-block analysis, using the formulas on p. 795 with λ^* and ν^* replaced by one.

Intra-block Variances and Standard Errors

a) *Varieties in same X Block*

$$\hat{V}_1' = \text{Var}\,\{v_{11}' - v_{12}'\} = \frac{2 \times 2.807}{5 \times 6}\left[6 + \frac{2}{3}\right] = 1.248$$

$$\text{S.E.}\,\{v_{11}' - v_{12}'\} = \sqrt{1.248} = 1.117$$

b) *Varieties in same Y Block*

$$\widehat{V}_2' = \text{Var}\left\{v_{11}' - v_{21}'\right\} = \frac{2 \times 2.807}{5 \times 6}\left[6 + \frac{3}{2}\right] = 1.404$$

$$\text{S.E.}\left\{v_{11}' - v_{21}'\right\} = \sqrt{1.404} = 1.185$$

c) *Varieties not in same Block*

$$\widehat{V}_3' = \text{Var}\left\{v_{11}' - v_{22}'\right\} = \frac{2 \times 2.807}{5 \times 6}\left[6 + \frac{2}{3} + \frac{3}{2}\right] = 1.528$$

$$\text{S.E.}\left\{v_{11}' - v_{22}'\right\} = \sqrt{1.528} = 1.236$$

d) *Average of all comparisons*

$$V_4' = \text{Av. Var} = \frac{2 \times 2.807}{5 \times 7}\left[6 + \frac{2}{3} + \frac{3}{2} + 1\right] = 1.470$$

$$\text{Av. S.E.} = \sqrt{1.470} = 1.212.$$

For most purposes the average standard error, 1.212, would be adequate.

The variances appropriate to the inter-block analysis are also derived from the formulas on p. 795, using the estimates $\hat{\lambda}^*$ and $\hat{\nu}^*$ of λ^* and ν^*.

Inter-block Variances and Standard Errors

a) *Varieties in same X Block*

$$\widehat{V}_1'' = \text{Var}\left\{v_{11}'' - v_{12}''\right\} = \frac{2 \times 2.807}{5 \times 6}\left[6 + \frac{2}{3}\,(0.3574)\right] = 1.167$$

$$\text{S.E.}\left\{v_{11}'' - v_{12}''\right\} = \sqrt{1.167} = 1.080$$

b) *Varieties in same Y Block*

$$\widehat{V}_2'' = \text{Var}\left\{v_{11}'' - v_{21}''\right\} = \frac{2 \times 2.807}{5 \times 6}\left[6 + \frac{3}{2}\,(0.2705)\right] = 1.199$$

$$\text{S.E.}\left\{v_{11}'' - v_{21}''\right\} = \sqrt{1.199} = 1.095$$

c) *Varieties not in same Block*

$$\widehat{V}_3'' = \text{Var}\left\{v_{11}'' - v_{22}''\right\} = \frac{2 \times 2.807}{5 \times 6}\left[6 + \frac{2}{3}\,(0.3574)\right.$$

$$+\frac{3}{2}(0.2705)\Big] = 1.243$$

$$\text{S.E. }\{v_{11}'' - v_{22}''\} = \sqrt{1.243} = 1.115$$

d) *Average of all comparisons*

$$\hat{V}_4'' = \text{Av. Var} = \frac{2 \times 2.807}{5 \times 7}\left[6 + \frac{2}{3}(0.3574) + \frac{3}{2}(0.2705) + 1\right]$$

$$= 1.226$$

$$\text{Av. S.E.} = \sqrt{1.226} = 1.107.$$

Again, for most purposes, the average standard error would be adequate. We see that the estimated average variance in the inter-block analysis is 17 per cent less than the estimated average variance in the intra-block analysis.

Due to the fact that $\hat{\lambda}^*$ and $\hat{\nu}^*$ are estimates rather than true values the above variance estimates tend to be a little too small, i.e. they have a negative bias. An approximate adjustment made to remove this bias changes the estimated standard errors in this case by less than one per cent [9].

E. *Efficiency Relative to Alternative Designs*

The formulas used to gauge relative efficiency are given on p. 798. In using them we will estimate σ_e^2 by the residual mean square, 2.807, and $k\sigma_B^2$ by the quantity $(b-e)R/(R-1) = (4.889-2.807)5/4 = 2.603$.

The estimated efficiency of this particular experiment using intra-block analysis relative to randomized blocks is

$$\frac{1.272}{1.470} = 0.865 \quad \text{or} \quad 86.5\%,$$

a loss of about 13 per cent. If the inter-block analysis is used, the relative efficiency becomes

$$\frac{1.272}{1.226} = 1.038 \quad \text{or} \quad 103.8\%,$$

a gain of not quite 4 per cent.

The efficiency relative to an equal sets design is very close to 100 per cent by either method of analysis. Using the intra-block analysis the efficiency is

$$\frac{6+3}{6+\frac{2}{3}+\frac{3}{2}+1} = 0.982 \quad \text{or} \quad 98.2\%.$$

Using the inter-block analysis we calculate an efficiency of 99.86 per cent.

The quintuple lattice does not exist for a 6×6 design, so there is no point in making that comparison for a lattice experiment with 36 varieties.

F. Conclusions on the Numerical Example

If the above findings were made relative to an actual experiment instead of our synthetic one, one might draw the following conclusions.

a) The lattice design did not appreciably improve the accuracy of the experiment relative to what might have been expected from a randomized blocks design.

b) The use of the inter-block analysis has saved the experiment from a considerable loss (13 per cent) relative to a randomized blocks design.

c) The loss of efficiency due to using unequal sets of replications was negligible.

[Note: In the two original experiments from which our data were taken, the apparent gains relative to randomized blocks (using inter-block analysis) were 11 per cent in the first experiment (from which we took four replications) and 3 per cent in the second (from which we took one replication).]

ACKNOWLEDGEMENT

The writer would like to express his thanks to Professor John W. Tukey for suggesting this problem.

BIBLIOGRAPHY

[1] Cochran, W. G., "Problems arising in the analysis of a series of similar experiments," *Journal of the Royal Statistical Society*, Supplement 4 (1937), 102–18.

[2] Cochran, W. G., and Cox, G. M., *Experimental Designs*, New York, John Wiley and Sons (1950).

[3] Cox, G. M., Eckhardt, R. C., and Cochran, W. G., "The analysis of lattice and triple lattice experiments in corn varietal tests," Iowa Agricultural Experiment Station Research Bulletin 281, 1940.

[4] Cramér, H., *Mathematical Methods of Statistics*, Princeton, Princeton University Press (1946), p. 116.

[5] David, F. N., and Johnson, N. L., "The effect of non-normality on the power

function of the *F*-test in the analysis of variance," *Biometrika*, 38 (1951), 43–57.

[6] Grundy, P. M., "The estimation of error in rectangular lattices," *Biometrics*, 6 (1950), 25–33.

[7] Harshbarger, B., "Rectangular lattices," Virginia Agricultural Experiment Station, Memoir 1, 1947.

[8] Kempthorne, O., and Federer, W. T., "The general theory of prime power lattice designs II," *Biometrics*, 4 (1948), 109–21.

[9] Meier, P., "Weighted means and lattice designs," Doctoral thesis, Princeton University, 1951.

[10] Nair, K. R., "Analysis of partially balanced incomplete block designs illustrated on the simple square and rectangular lattices," *Biometrics*, 8 (1952), 122–55.

[11] Yates, F., "A new method of arranging variety trials involving a large number of varieties," *Journal of Agricultural Science*, 26 (1936), 424–55.

[12] Yates, F., "The recovery of inter-block information in variety trials arranged in three-dimensional lattices," *Annals of Eugenics*, 9 (1939), 136–56.

ON THE PRESENTATION OF THE RESULTS OF SAMPLE SURVEYS AS LEGAL EVIDENCE*

W. EDWARDS DEMING
New York University

PURPOSE OF THIS PAPER

THE purpose here is to view some of the problems that confront the statistician when he presents the results of a sample survey as legal evidence. One particular point is that the statistician, if he is to make his work useful, must distinguish between (a) what he as a statistician may say about the *precision* of the results of his survey, and (b) what an expert in the substantive field may conclude about the *usefulness* of the results. The statistician can testify only to the former, and possibly also about the variance between investigators, and between different methods, if he measured these differences. As a secondary purpose, we shall enquire into the meaning of a standard error, and its relation to a complete count and to the usefulness of the results—a point that is often overlooked, not only in testimony but in statistical reports.

I have no magic nor all the answers to all the questions and difficulties that the statistician will encounter when he presents results as evidence. It is possible, however, to share some experiences with colleagues in this increasingly important role of statistical surveys; to acquaint them with some of the kinds of problems that may arise; and to suggest some general principles that will help the statistician to make his work more useful than it would be otherwise.

At the outset I may explain that this paper will deal only with probability samples. The defence of any other kind of sample is hardly a problem for a statistician anyhow, but rather for the substantive expert who may have enough knowledge of the material and of its variability to feel that he can testify one way or another with respect to the interpretation of the results of a judgment sample.

A statistical survey, formulated and carried out by the dictates of the theory of probability, is to the statistician an exciting and remarkable achievement. It produces man's best empirical knowledge, and it provides an objective measure of the amount of knowledge in the survey. The precision desired can be aimed at and hit pretty accurately by planning *in advance* with the aid of the theory of probability and with bits of knowledge with respect to certain proportions, means, correlations, variances, and other statistical measures of the sampling units in the frame. Then,

* Presented at a meeting of the American Statistical Association in Chicago, 29 December 1952.

after the survey is completed, the precision that was actually reached is calculable and expressible in an international standard of measure (the standard error) from the results of the survey itself. This final measure of the precision is objective, and is not a matter of opinion. It is not biased by incorrect assumptions that went into the planning.

The statistician when he presents his results as legal evidence finds himself nevertheless at an uncomfortable disadvantage. He is usually talking to scholars, but not to fellow statisticians, nor indeed to other scientists, nor relating the results of his research to a trusting client or sponsor. He is teaching, but the techniques of the class room will not necessarily be the best ones for the presentation of evidence.

Scholars in other disciplines are not all acquainted with the achievements of probability sampling, yet the statistician must somehow explain his methods to them. Some of the people that he must deal with in legal evidence know sampling only as a failure to predict an election; they know not the distinction between (a) the standard error of sampling, (b) the errors common to complete counts and to samples, and (c) the error of a prediction. Other people think of sampling as a selection by judgment, carried out by someone who has established a reputation by a run of successes in the past. To still others, sampling is a desperate risk, a hazardous aimless random drawing of areas or of other elements to which anything may happen, and concerning which nothing can really ever be known except by comparison with a complete count, for which a sample is only a substitute to save time and money.

In my own experience, a man questioned the existence of the theory for estimating the variance of a mean, originated by Gauss 120 years ago, and now used all over the world. I was once accused of "pyramiding" my results (whatever that is), because I took the average of the averages of my 10 subsamples for an estimate of the whole, wherefore my standard error "must be viewed with some doubt."

DIRECT TESTIMONY: CROSS-EXAMINATION

There is first of all direct testimony, wherein the statistician presents his results, after careful preparation in advance. He will usually simply read his direct testimony into the record from typed copy. Direct testimony may take the form of questions and answers, the questions being read by the lawyer who has engaged the statistician. The questions should be framed so that they display the results of the survey in the form of valid statistical inferences. The questions must not sound as if they were digging for particular answers, even though both sides in the case know full well that the statistician is reading from prepared copy, and that the

answers are exactly what the statistician believes to be essential to his methods and to his results, regardless of the questions.

In the preparation of testimony, the lawyer who engaged the statistician will not try to influence the content of the statistician's statements. He will try to help the statistician to state his procedures and his inferences so that they will be clear. The inferences must be only what the statistician can support, as a scientist seeking truth. To bring out the truth in a scientific inference, one must not only state what he believes to be true, but he must say it so that his listeners will understand what he means, and not think that he has said something that he did not mean. A good lawyer can help immeasurably in achieving this aim.

In giving evidence, the statistician is not fighting a case for either side. He is an expert witness, and he should appear as a professional man, with the sole aim of presenting the truth. This means that he must tell to the best of his ability what the figures mean. He should describe in full any difficulties that he encountered, and their possible limitations on the interpretations of his data.

In some courts one may not read prepared testimony, in which case one can only prepare to present his testimony without the aid of his typed copy. He will of course still be able to present tables and charts, called exhibits.

Usually during or immediately following direct testimony the opposing side asks only questions that will clear up simple failure to recognize technical terms, or to clarify some events with respect to their sequence in time. Questions that may bring out flaws in the testimony, they will usually reserve for further study, following which they will call the statistician to the stand for cross-examination. Here the questions are often well-prepared in advance, but the statistician must answer *ex tempore.* Here the statistician may find himself very uncomfortable to find that statements and interpretations that he thought were clear and objective are now misunderstood and misinterpreted, and his statistical principles challenged.

When cross-examination comes, no matter what question comes, relevant or irrelevant, do the best that you can with it. Be cautious to stay within the field of competence that you have testified to in your qualifications (vide infra). Groundwork in your direct testimony, in an attempt to give clear explanations of your procedures, of the statistical interpretation of your results and of their standard errors, will help to keep the cross-examination on the track and to bring out the inherent scientific truth contained in your survey.

To present the results of a survey in a case where millions of dollars are involved, to ears unfamiliar with the power of modern statistical practice, is an experience that purifies the statistician's thinking. Sometimes the listeners are glad to accept the results of a good survey, and to learn something about modern survey-methods. At other times, they will declare that the statistician's methods are new and untried, that his results are therefore not acceptable evidence; that his sample was too small; or finally, foresooth, that he has not explained the entire theory of sampling so that everyone can understand exactly what he did and why, and that there is therefore no basis by which to judge whether his results have any meaning.

ESSENTIAL INGREDIENTS OF THE DIRECT TESTIMONY

The statistician's statement of his qualifications, which usually comes in the first part of his direct testimony, is important. It is evidence by which the examiner or judge may decide, if the question arises, whether the statistician is qualified. It should therefore contain a full account of the statistician's education and relevant experience.

He may then present the purpose of the survey (an example of an assignment will occur later), what he endeavored to do, the methods that he prescribed, the basis for these methods, the system and the observations by which he satisfied himself that the procedures that he prescribed were understood and followed rigidly and faithfully; finally, the results and their standard errors and their interpretation; also the possible effects of any biases inherent in the procedure, and the possible effects of any difficulties encountered. All these points will go into the direct testimony.

He should tell in simple words what the procedures actually were. He should limit theory to a few simple and well-established principles that illustrate the sampling procedures and the interpretation of the results. The truth and the whole truth means clarity, so that anyone may judge whether your results and your interpretation of the standard error are what you say they are. You can not hope to give a whole course in the theory of sampling, but you can make your procedures and their validity clear without doing so. The most convincing argument concerning your procedures and of your interpretations is that they conform to established international standards, and that they are used in a wide variety of experience. In this connection the document written by the United Nations Sub-Commission on Statistical Sampling entitled, "The presentation of sampling survey results" (UN Series C, No. 1, 1950) is of assistance; likewise the "Manual on the Quality Control of Materials" (1951)

and other recommended and standard practices of the American Society for Testing Materials, many of which have been adopted as standards in other parts of the world.

A formula can cause trouble unless you explain pretty expertly how you used it. If in direct testimony you say that you used a formula to calculate in advance the size of sample required, when the fact is that you made a rough mental calculation and tempered it with judgment, or that you made the calculation years ago for similar work, and really did not make a fresh detailed calculation for this job, or if you did make one and then modified the answer to allow for some possible additional variance not fully represented in the formula, or to allow for some possible heavy additional cost of inspection or of interviewing because of probable bad weather in February, and if now upon cross-examination when people start asking questions you can not get exactly the same sample-size out of your formula as you actually used, you may find yourself very uncomfortable. The trouble is that people not accustomed to formulas will not understand how one uses theory.

If you say that a certain constant in your formula for the required sample-size represents your advance estimate of the variability of the material that you sampled, someone may accuse you of prejudging the answer. The fact is, however, that this advance estimate does not invalidate in the slightest the standard error calculated from the results, nor cause any bias in the procedure. You must make this clear in your direct testimony.

In practice sample-sizes are based on both theory and experience, even though you do not make a fresh calculation for every sample-design. Theory is part of your experience. Without theory, experience has no meaning. Theory and experience together produce scientific advances. All this can be made clear, I believe.

I proceed now to describe some of the other problems of exposition that have arisen, and to offer some suggestions toward meeting them.

IMPLICIT FAITH IN THE COMPLETE COVERAGE, AND IN THE 10 PER CENT SAMPLE

A complete coverage, no matter how carried out, and even though it is incomplete (as complete counts too often are), has weight in evidence. A sample, unless it is a 10 per cent sample, has two strikes against it to start with. People who are not statisticians assume that the sheer size of a complete coverage will somehow cover up its incompleteness and the flaws in the method of measurement or in the interviewing. They believe that a judgment sample, if it is big enough, will do the same; and that it

will in addition overcome biases of the unknown probabilities of selection.

A 10 per cent sample has almost equal standing with a complete count —maybe even better than a 15 per cent sample. Why, or what 10 per cent, is hardly ever questioned, even by experts in quantitative subject-matter.

The statistician, in the explanation of his sampling procedure, faces such preconceived ideas. The precision of a small sample, selected and estimated by an efficient probability procedure, will require justification. It is a fact that the aerial plant in a sample of 1000 to 1500 telephone poles will provide all the precision that one can use for the estimation of the average over-all physical condition of the entire aerial plant which might be worth $200,000,000. But without very careful preparation to dispel preconceived ideas about complete counts and 10 per cent samples, the statistician must be prepared to face an objection on the ground that a sample of only 1 part in 1000 is not admissible as evidence. The man who objects may, without knowing it, own stock in a woolen mill that purchases a million pounds of wool and pays duty on it on the basis of a sample that weighs from 60 to 100 ounces.

The troubles that people have in understanding the power of a small sample are often tied up with failure to understand that it is the absolute size (n) of the sample, and not its proportion (n/N) to the whole, which determines the standard error of the result. The statistician must be prepared to meet the man who thinks that to reach a prescribed precision in an estimated average rent, for example, a sample of dwelling units from a big city must be bigger than the sample from a small city, because the big city is bigger.

With careful preparation, you can dispel such misunderstandings in an entertaining way, and in simple language. You can explain with black and white beans the statistical principles used, and why it is that the standard error of a sample is in practice hardly influenced at all by the size of the lot that it was drawn from. You can portray vividly how a pint jar of dried beans scooped up from a larger mixture of black and white beans will provide an estimate of the proportion black in the mixture; and that a sample of less than a pint would probably be sufficient. You may then observe, and your listeners will agree, that the mixture could as well be a carload of beans as a bushel of beans: the sample provides as good an estimate of the proportion black for the carload as it does for the bushel, *provided* that in both cases the mixture is thoroughly mixed (an illustration borrowed from testimony presented by Professor John W. Tukey). In practice we accomplish thorough mixing with the use of a table of random numbers—a tool indispensable today in science.

The nigh total failure of the size of a lot to have any influence on the standard error of a random sample drawn therefrom is illustrated by charts in Eugene L. Grant's book, *Statistical Quality Control* (McGraw-Hill, 1946), page 345. Incidentally, such citations will often help the statistician's listeners to appreciate the fact that his methods are in universal use. One may usefully refer to the ever-expanding dependence of all kinds of scientific, industrial, agricultural, and medical research on statistical theory; the use of statistical methods to attain extreme precision in industrial production; the necessity for proper statistical design in the comparison of two industrial processes, machines, or medical treatments, the growing reliance, in many parts of the world, on probability samples in social and economic studies that are to guide important decisions.

If you succeed in making your explanation clear, you will help your listeners to appreciate the contribution of modern statistical principles and techniques to scientific truth. They may be grateful, in the long run.

Complex terms, flourished too freely, may alienate your listeners. Rely on patience, truth, and simple language. You can not afford to lose the attention of the examiner or judge; he is in position to protect truth and accuracy of statement. In cross-examination keep him on your side by your fairness and willingness to try to clear up any questions concerned with your sample.

PRECISION, ACCURACY, AND STANDARD ERROR

Two concepts that are important to make clear in any presentation are precision and accuracy. Most statisticians probably think that they know what these words mean. I must confess that experience under the fire of cross-examination taught me some new angles to their meaning, and taught me the importance of explaining in advance the limitations of a standard error.

Precision is expressible by an international standard, viz., the standard error. It measures the average of the differences between a complete coverage and a long series of estimates formed from samples drawn from this complete coverage by a particular procedure of drawing and processed by a particular estimating formula.

Great precision or a small standard error attached to an estimate does not mean that this estimate is necessarily highly accurate or useful. It does mean that the results of a complete coverage would have been the same within a very narrow margin of difference, had the complete coverage been carried out with the same investigators, sharing the load proportionately, and with the same care as they expended on the samples.

The so-called "expected value" of a sampling procedure (which of course includes the formula for the estimate) is the same as the result of an attempted complete coverage of the same frame that the samples are to be drawn from (except for a possible bias in the formula, for which an upper and innocuous limit will be known). Both the complete coverage and the sample are subject to the same uncertainties and errors, such as inadequate supervision, nonresponse, wrong information, missing information, failure of workers to cover their whole assignments, and to find all the people or all the items. The only difference is that the sample has sampling error, which is the one error that we are best able to govern and to measure. The statistician measures the uncertainty introduced by sampling. The substantive expert judges whether the same operations would give accurate and useful information if applied to the entire frame.

The statistician will have drawn up the statistical procedures for the survey (the design of the sample, the instructions for drawing it, the instructions for tabulating the results and for computing the estimates and their standard errors). During the progress of the work, he should be on hand as often and as long as necessary to know that the company that retained him is following his instructions meticulously. He is then in a position to defend the validity of the standard error. If at any time he is not satisfied with the performance of the workers, it is better for him to terminate at once his relationship with the client. He should be sure that this responsibility is clear beforehand.

A statistician will occasionally be called upon to give his opinion in regard to procedures that another statistician has drawn up and testified to, or to give his interpretation of the results, including the standard error. After he has a chance to examine the procedures, he may testify, if he agrees, that they are one of many possible probability designs, and that IF they were followed meticulously, the results and the standard errors have certain interpretations, which he may give if called upon to do so. He may require, before he testifies, that certain calculations be carried out, to help him to examine the magnitudes of any biases that he may suspect. He may require calculations of skewness, if he suspects that the estimate of the standard error is not sufficiently firm. The results of these investigations will guide his conclusions and his testimony concerning the precision of the results of the survey. He must not be satisfied to testify to what he knows; he must explain how certain aspects of the survey that he had no opportunity to examine could possibly affect the results.

Even with familiarity with the job, and no matter how satisfied the

statistician may be with the execution thereof, he can still not testify to the inherent usefulness of the result. Unfortunately, he has no standard error of the usefulness of a result. Testimony on the usefulness of the results will be left to the substantive expert—the engineer, the chemist, the physician, the population expert, the agricultural expert. The usefulness of a result is not a problem of sampling; it deals rather with the method of measurement and with reasons why the method used will produce data that will satisfy a particular need. The method would be the same whether the survey were a complete coverage or a sample.

In cross-examination the opposition may tempt the statistician beyond the sphere of his competence. The statistician must try to answer all questions politely and simply, yet he must stay within the limitations of his own ability and of the standard error. He certainly has a right to say he does not know the answer to a question that is beyond his competence and beyond his direct knowledge.

Although he can not testify to the inherent usefulness of the result, the statistician can certainly make it clear that he would not have associated himself with the study had he not been sure in advance that it would be executed rigidly in conformance with his specifications, and that the methods of inspection, interviewing, and questioning, although beyond his qualifications, would be satisfactory and produce useful data. He may do this without professing to be an expert in the subject-matter, as he may declare that he has confidence in Mr. So and So (expert in the subject-matter), who has testified, or will, concerning these things.

This division of responsibility between the statistician and the expert in the subject-matter should not be difficult to explain, but it is easy to forget to do it; and still easier later on in cross-examination to be lured across the border into the subject-matter and into trouble.

The following excerpt represents a statistician's attempt in direct testimony to state what his job was; to put a limitation on his assignment, and hence on what he could testify to in cross-examination. The case involved the use of samples of items of telephone plant, the aim being to obtain a figure for the average over-all per cent physical condition of the entire property.

> *Q.*[1] Doctor, for what purpose were you engaged by the Illinois Bell Telephone Company?
>
> *A.* I was told that this company proposed to make a survey to determine the physical condition of its plant. They asked me to prescribe statistical procedures by which to select samples of items of plant for inspection,

[1] The Illinois Commerce Commission, Docket No. 39126, 1951, and Docket No. 41606, 1954: The Illinois Bell Telephone Company in the matter of the proposed advance in rates. The passage printed here is testimony prepared in advance, and is not necessarily the same word for word in the record.

such as poles, wire, cable, telephones, relays, central office equipment. The samples must determine within narrow limits of precision what result would be obtained for the average over-all per cent condition by a complete 100% inspection of all the items in all the classes of plant that were to be inspected, with the same inspectors, and with the same care as was exercised on the samples, were such a thing possible.

This assignment carried with it the responsibility for prescribing the procedures for summarizing the results for each class of plant, once the code-values assigned by the inspectors were translated into percentages, and for combining the per cent conditions of the several classes of plant into the over-all average per cent conditions of all the classes that were to be inspected. A necessary part of the assignment was to provide procedures by which to calculate the standard error of the precision of the result obtained for the over-all average per cent condition.

My assignment did not include the responsibility for the procedures for inspecting any item, nor for the numerical values that translated the inspectors' codes into percentages. Neither had I any responsibility for determining the weights of the various classes of property. These problems are the same whether one uses sampling or not. These phases of the work have been described by Mr. Coxe (General Staff Engineer).

Q. Does Company Exhibit No. 112 (Sampling Procedures for Drawing the Items of Property for Field Inspection) contain the procedures that you prescribed?

A. Yes sir, it does.

Later on came the following explanation of the standard error of the result:

Q. What is your interpretation of the standard error of this study?

A. The sampling precision of this study is expressed by the over-all standard error, which turned out to be .19 per cent. This standard error is not a matter of opinion nor of expert judgment, but is objective, as it is calculated by the laws of probability from the results themselves. The interpretation of this standard error is simple: I may say with a high degree of assurance that the maximum uncertainty that one may attach to the over-all per cent condition because of the introduction of sampling, can not be, at the outside, more than three times the standard error. In other words, any uncertainty in the figure 74.5% (the final result) which can be attributed to the fact that the company used samples instead of a complete and total inspection of every item, with the same care as was exercised on the samples, were such a thing possible, can not exceed .57 per cent.

THE PERMANENCE OF THE STANDARD ERROR CONTRASTED WITH THE TEMPORAL CHARACTER OF ACCURACY AND USEFULNESS

In a probability sample (the only type of survey to be considered here) the precision is calculable from the results, as I mentioned in the opening paragraphs. In practice, the size of the sample will be sufficient to provide

a firm estimate of the standard error. This was the case in the excerpt above. One may say with a high degree of assurance that in a long series of repetitions of this sampling procedure, only about 2.3 per cent of the results would fall 2 standard errors above the result of the complete coverage, and about 2.3 per cent of the results would fall 2 standard errors below. Practically none of the long series would fall beyond 3 standard errors either way.

It is another thing, however, to say whether the complete coverage, were it possible, would produce useful information. The inherent accuracy of the method of measurement (the interviewing, the questionnaire, the method of inspection), and the usefulness of the information, whether obtained by a complete coverage or by a sample, is a matter for the substantive expert to testify to, as explained earlier.

The main difference between a sample and a complete count is that the sample possesses an error of sampling. The statistician testifies in regard to this. A sufficient degree of precision is necessary for the usefulness of sample results, but it does not guarantee their usefulness.

The inherent accuracy and usefulness of the procedures of measurement will change from time to time as the substantive experts develop new concepts of the kind of information that they require to solve new and changing problems. Anyone who has followed the changing concepts of the characteristics of the labor force, or the changing concepts of a farm, or of family-budget studies, or the changing concepts of the desirable characteristics of fibres and of textiles, will know that no definitions or methods of measurements stay fixed.

In contrast, the standard error of a procedure of sampling remains fixed with time; likewise the interpretation of the standard error. The validity of the standard error does not depend on the economy or cleverness of the design of the sample. It depends only on careful execution and on the rigid use of probability methods in accordance with some prescribed statistical plan. For this reason, the standard error of a sampling procedure and its interpretation, remain valid, even though new advances in theory point the way to more economical sampling procedures by which to obtain the same standard error.

REFERENCES

There is apparently no previous literature that deals with the presentation of modern statistical procedures and their results in legal evidence. Fortunately, however, sampling has received attention from the legal standpoint in a paper by Frank R. Kennedy, who supplied copious re-

marks and references to cases in which samples of one kind or another were offered in evidence.[2]

In conclusion, it is a pleasure to acknowledge aid from a number of friends, chiefly from Mr. Melvin F. Wingersky, Attorney at Law, and a member of this Association. This acknowledgment is of special interest, because he cross-examined me on several occasions with great vigor. Valuable help came also from Mr. Harlow A. Coxe, General Staff Engineer of the Illinois Bell Telephone Company; also from Mr. Howard L. Jones, statistician, and from Mr. Gordon Winks, General attorney, all with the same company. Finally, I have had the benefit of a long association and many conversations on this subject with Professors John W. Tukey and Frederick F. Stephan of Princeton.

---

[2] Frank R. Kennedy, "Some legal aspects of sampling," Industrial Quality Control, Vol. vii (January and March 1951).

ACCURACY OF AGE REPORTING IN THE 1950 UNITED STATES CENSUS

ROBERT J. MYERS
Social Security Administration

ONE common human error is to round figures even though precise results might be desired or requested. This is particularly evident in census returns where the age in integral years is sought. Particularly, does this arise for ages ending with the digit 0 and to a lesser extent frequently with digits 2, 5, and 8. This paper will investigate the extent of preference for certain digits of age in the 1950 United States census and will indicate the extent of improvement that has occurred since earlier censuses, as well as giving certain summary data for several other countries. The analysis will be carried out using the "blended" method.[1]

DESCRIPTION OF METHOD OF ANALYSIS

One method for showing the degree of preference for certain digits of age consists of starting at a given age, say 20, and adding up the population for all ages ending in 0, all ending in 1, etc. Then the population at each digit is expressed as a percentage of the total population; any considerable deviation from 10 per cent would be taken as indication of bias in age reporting for that particular digit. This procedure, however, does not yield truly valid results since it is not proper to simply add the overall populations at each digit starting at a particular age because then the "leading" digits naturally occur more frequently among the persons counted than the "following" ones.

The "blended" method overcomes this objection by allowing each digit in turn to be the "initial" one. The ten separate results are then summed, and a percentage distribution by digit is computed. The justification for this method is largely empirical, based on general reasoning and logic, with the further point that it produces proper results for smooth, life table data (i.e., shows no digit preference).

As an example of how the "blended" method operates, when the count is started at age 20, the population considered at unit digit 0 is the sum of those at ages 20, 30, 40, etc. For the nine cases when the count is started successively at ages 21 to 29, the population considered at digit 0 begins with that at age 30 instead of age 20. Correspondingly, as to the population at digit 1, when the count is started in turn at

[1] See Robert J. Myers, "Errors and bias in the reporting of ages in census data," *Transactions of the Actuarial Society of America*, XLI (1940). Reproduced in *Handbook of Statistical Methods for Demographers*, Bureau of the Census, 1951.

ages 20 and 21, included are ages 21, 31, 41, etc., while when the count is started at any of ages 22 to 29, included are ages 31, 41, etc.[2]

The result of these calculations then is a percentage distribution of the population at each of the 10 digits. If no heaping were present, each figure would be very close to 10 per cent. Conversely, any sizable deviation from 10 per cent indicates the presence of such inaccuracy. A relative index of the amount of preference of age for any census distribution can be obtained by summing up the absolute deviations from 10 per cent in each case.

Bachi[2] has suggested a somewhat preferable index, which amounts to half the previously described index[3] and which will hereafter be used as the index of heaping. This index has a certain significance since, as Bachi says, it "estimates the proportion of persons in the population who return their ages with an inaccurate unit digit and thus has the advantage of being more easily understood." Bachi goes on to take the extreme case where all people report the same unit digit in which case the index would be 90 per cent indicating that 90 per cent of the people returned inaccurate unit digits.

In actuality, Bachi's index more properly indicates the *minimum* proportion of persons returning their ages with an inaccurate unit digit since certain errors may be self-cancelling. Thus, taking the common case where digit 0 is over-reported, there may be some persons truly having an age ending in 0 who report some other age; these persons are, of course, far more than offset by those who inaccurately report themselves at an age ending in 0. In the extreme case, of course, there might be 10 per cent of the persons reported at each of the 10 digits of age, which would yield an index of 0; yet it is theoretically conceivable that every person has returned his age with an inaccurate unit digit, but by chance there has been complete offsetting. At any rate, however, the use of the index as developed by Bachi seems preferable because it does have a certain real meaning.

ANALYSIS OF U. S. CENSUS DATA

Table 1 shows the preference for digits of age in the total United States population for various censuses. Considerable heaping at digits 0 and 5 occurred in the past, although there has been much improvement in the past 70 years. Thus, in 1880, digit 0 showed a relative

[2] A very similar method of analysis, which in practice produces very much the same results, was developed independently by Roberto Bachi in "Measurement of the tendency to round off age returns," *Proceedings of the International Statistical Congress*, Rome, 1953.

[3] Or in other words, is based on only the preferred, or overstated, digits (or conversely, only on the disliked, or understated, digits).

excess of 68 per cent over the normal proportion of 10 per cent, thoeretically to be expected, whereas by 1950 this excess amounted to only 12 per cent. The heaping at digit 5, which in 1880 was half as large as that at digit 0, has by 1950 virtually vanished. Throughout the period a slight heaping at digit 8 has apparently been present. The greatest understatement has occurred for digit 1, thus indicating that the heaping at digit 0 seems to be due to digit 1 rather than digit 9. The index of preference has decreased steadily from more than 10 in 1880 to almost 2 in 1950.

TABLE 1

PREFERENCE FOR DIGITS OF AGE IN THE TOTAL CONTINENTAL UNITED STATES POPULATION FOR VARIOUS CENSUSES, POPULATION AT EACH DIGIT OF AGE AS PER CENT OF TOTAL POPULATION[a]

Digit of Age	Census							
	1880	1890	1900	1910	1920	1930	1940	1950[b]
0	16.8	15.1	13.2	13.2	12.4	12.3	11.6	11.2
1	6.7	7.4	8.3	7.7	8.0	8.0	8.5	8.9
2	9.4	9.7	9.8	10.2	10.2	10.3	10.4	10.2
3	8.6	9.1	9.3	9.2	9.4	9.4	9.6	9.7
4	8.8	9.0	9.5	9.4	9.4	9.6	9.7	9.7
5	13.4	12.3	11.3	11.5	11.3	11.2	10.7	10.6
6	9.4	9.6	9.4	9.6	9.7	9.6	9.6	9.8
7	8.5	8.9	9.3	9.1	9.4	9.3	9.6	9.7
8	10.2	10.4	10.2	10.7	10.6	10.5	10.3	10.2
9	8.2	8.5	9.7	9.4	9.6	9.8	10.0	10.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Index[c]	10.4	7.8	4.7	5.6	4.5	4.3	3.0	2.2

[a] The method of analysis is the "blended" method as described in the text, using starting ages 23 to 32 and ending at age 99 in all cases. These percentages, in effect, relate the reported population at each digit of age to the "true" population.

[b] Based on 20 per cent sample data.

[c] The index is one-half the sum of the deviations from 10.0 per cent, each taken without regard to sign. These indexes, in effect, indicate the minimum net proportion of persons who return their ages with an inaccurate unit digit.

Table 2 indicates the variation in preference for digits of age by sex, race, and nativity for the 1950 census. For native-born whites, there is very little preference for any particular digit of age although there is a small heaping at digit 0, and to some extent at digit 5. The index of preference is at the very low level of 1.3 for native-born white men although being somewhat higher for women of the corresponding group.

The foreign-born whites show greater inaccuracy in the reporting of ages than native-born whites, with the index of preference being about twice as high, while for nonwhites, the index is about 4 times as high as for native-born whites.

TABLE 2

PREFERENCE FOR DIGITS OF AGE BY RACE AND SEX IN 1950 CENSUS OF CONTINENTAL UNITED STATES POPULATION AT EACH DIGIT OF AGE AS PER CENT OF TOTAL POPULATION[a]

Digit of Age	Men			Women		
	Native-born White	Foreign-born White	Non-white	Native-born White	Foreign-born White	Non-white
0	10.6	11.4	13.0	11.0	12.2	13.4
1	9.2	8.4	7.5	9.0	8.0	7.1
2	10.2	10.3	9.9	10.2	10.1	9.8
3	9.8	9.7	8.9	9.8	9.5	8.7
4	9.9	9.8	9.3	9.7	9.6	9.0
5	10.3	11.1	11.5	10.5	11.4	11.5
6	9.8	9.6	9.4	9.8	9.5	9.4
7	9.9	9.7	9.4	9.7	9.5	9.4
8	10.0	9.9	10.5	10.2	10.3	10.9
9	10.1	10.0	10.6	10.0	9.9	10.7
Total	99.8	99.9	100.0	99.9	100.0	99.9
Index[b]	1.3	2.8	5.6	2.0	4.0	6.6

[a] The method of analysis is the "blended" method as described in the text, using starting ages 23 to 32 and ending at age 99 in all cases. These percentages, in effect, relate the reported population at each digit of age to the "true" population. Based on 20 per cent sample data.

[b] The index is one-half the sum of the deviations from 10.0 per cent, each taken without regard to sign. These indexes, in effect, indicate the minimum net proportion of persons who return their ages with an inaccurate unit digit.

There is considerable evidence of greater accuracy of reporting in the 1950 census since the indices for all categories are considerably lower than in previous censuses (see Table 3). In each category and for each census, the index of preference is lower for men than for women, with the relative differential being about ⅓ for native-born whites although generally somewhat less than this for foreign-born whites and non-whites.

ANALYSIS OF CENSUS DATA FOR OTHER COUNTRIES

Some indication of the relative accuracy of the reporting of digits of age in the 1950 United States census may be obtained by considering

TABLE 3

INDICES SHOWING PREFERENCE FOR DIGITS OF AGE, BY RACE AND SEX, CONTINENTAL UNITED STATES, 1930, 1940, AND 1950 CENSUSES[a]

Sex and Race	Census		
	1930	1940	1950[b]
Men, Native-born White	2.8	2.1	1.3
Men, Foreign-born White	5.5	3.7	2.8
Men, Non-white	12.0	8.1	5.6
Women, Native-born White	3.4	2.8	2.0
Women, Foreign-born White	6.0	4.8	4.0
Women, Non-white	12.8	8.6	6.6

[a] The method of analysis is the "blended" method as described in the text, using starting ages 23 to 32 (except for white population in 1940, for which ages 35 to 44 were, of necessity, used) and ending at age 99 in all cases. The index, in effect, indicates the minimum net proportion of persons who return their ages with an inaccurate unit digit.

[b] Based on 20 per cent sample data.

the indices of preference in recent censuses in other countries where it would be expected that, because of a high degree of literacy, good reporting would be obtained. Data by single years of age sufficient to make such an analysis are available for Australia (1947), Canada (1951), and Great Britain (1951), with the results being as follows:

Country	Index of Preference	
	Men	Women
Australia	1.2	1.4
Canada	1.3	1.6
Great Britain	1.4	1.1
United States, Total	1.9	2.4
United States, White	1.5	2.2

As contrasted with the other three countries, the index of preference for the United States is significantly higher, especially for women. However, when only the white population of the United States is considered, the index for men compares quite favorably with those for the other countries, but for women the United States index is significantly higher. In fact, it is pertinent to note that for the other countries, there is relatively little difference in the accuracy of reporting of ages as

between men and women, whereas for the United States women definitely do not report as accurately as men.

In both Canada and Great Britain, the only significant evidence of heaping is for digit 0, but this is of relatively minor significance representing an overstatement of at most 10 per cent relatively. For Australia, the situation is somewhat different since there is only a slight indication of heaping at digit 0; in fact, there is evidence of a certain amount of heaping at digit 7. This peculiarity possibly arises because the census was taken in 1947, and the question on age was framed so as to ask for year of birth. Accordingly, a very sizable number of persons reported the "round" year, 1900 (the number shown at age 47 being 10 per cent greater than the average at ages 46 and 48).

SUMMARY AND CONCLUSIONS

The accuracy of the reporting of ages in the 1950 United States census has been in accord with the trend of steady improvement prevailing over the last 70 years. For certain groups, especially native-born white males, age reporting now, at least insofar as preference for digits of age is concerned, has reached almost as great accuracy as can ever by expected. There is, however, significant room for improvement in the nonwhite population. Furthermore, the reporting of ages by women is significantly less accurate than for men despite the fact that in various foreign countries there is little difference between the sexes as to accuracy of age reporting.

VALIDATION OF MORBIDITY SURVEY DATA BY COMPARISON WITH HOSPITAL RECORDS*

NEDRA B. BELLOC

California State Department of Public Health†

HOUSEHOLD sample surveys are being used increasingly[1] to obtain information on the status of the health of the population. Physicians, and others who are to use many of the data so collected, are likely to raise the question: When lay interviewers, no matter how carefully trained, question lay respondents on a subject as complex as illness, are the results sufficiently accurate to justify the continued use of this method of measuring morbidity?

Validation of the measures of morbidity by comparison with an independent criterion can be done only if the person who reported illness received some medical service. Reports of absence from school or work are not necessarily proof of illness. This means that a large proportion of illnesses reported in household surveys are not subject to verification, since they are not medically attended.

In almost all of the earlier illness surveys some attention was paid to the assessment of the accuracy of the diagnostic information obtained [7, 11, 13]. In some cases the effort was directed at "improving" the diagnoses given by respondents [16, 17] by the substitution of medical reports for those given in the surveys. Only in the National Health Survey [8, 14] were sufficient data presented to enable an evaluation of the extent of agreement between the family's and the physician's diagnoses. In that study, however, as in several of the others [2, 5, 6, 10], the diagnoses as given by the family were submitted to the physician for confirmation or change, creating an obvious prejudice in favor of agreement of the diagnosis.

The method used in these surveys has been the checking of a survey report against a corresponding physician record or hospital record. The degree of agreement has been described as a percentage of records in which the diagnosis agreed out of those which were checked. Another

* This investigation was supported by a research grant (RG-1792) from the National Institutes of Health, U. S. Public Health Service. Presented at the 113th Annual Meeting of the American Statistical Association, Washington, D. C., December 29, 1953

† Acknowledgment is made of the planning and preliminary work on this study which were done by Arthur Weissman.

1 Such studies as the Pittsburgh Arsenal Health District Studies, Canadian Sickness Survey, surveys sponsored by the Commission on Chronic Illness in Hunterdon County and Baltimore, the Special Research Project of the Health Insurance Plan of Greater New York, as well as the California Morbidity Research Project.

aspect of the validity of survey data, that of completeness of reporting of the fact of illness, has been virtually ignored.

It is to be noted that, while it is valuable in some ways, the matching of individual records with criterion sources does not give a statistical measure of bias. Individual reports might differ considerably from the corresponding reports in the criterion source, and yet the errors might compensate in such a way that the over-all description of morbidity given by the two sources could be identical. The statistician is interested in a test of validity which will enable him to answer the question, "Does the measure of morbidity obtained by a household survey differ significantly from that obtained by reference to medical records?"

Comparison with "criterion sources" can be made without the implication that these criteria are more accurate than the survey. In some cases, it is quite probable that the household survey reports are more complete than the medical records. If the completeness of reporting in the household survey is to be measured, however, it is necessary to have a criterion source which is in itself extremely accurate. Otherwise, the household survey will produce a large number of over-reports solely because of the inadequacy of the criterion.

This paper will present some of the results of the validation checks with records of hospitalization which were done in a survey undertaken by the California State Department of Public Health in San Jose in the spring of 1952 [1, 18]. The method used was to collect abstracts of records from the hospitals serving the City of San Jose for all persons resident in the city, and then to locate in this file of abstracts the records of hospitalization for all persons in the household sample survey. The two sets of reports thus obtained, from the household survey and from hospital records, form the bases for the computation of various measures. The net differences[2] between these two sets of statistics are examined in this report.[3]

The measurement of hospitalized illness by securing records from hospitals for a sample of the population required intensive work with the hospitals, and would not be practical except in a limited area such as San Jose. This study could not have been done without the patience, interest, and whole-hearted cooperation of administrative and medical record personnel in the participating hospitals.

---

[2] Marks and Mauldin [12] have classified errors in surveys as of three types: sampling, response, and processing. Since our survey reports are processed data, the error being studied here, while primarily response error, includes errors of processing.

[3] Since there is some interest in the extent of agreement on "matched cases," the method used in other surveys for validation, footnote references are made to analyses made by this method also.

Data on Hospitalized Illness Obtained from Hospital Records

For the seven months prior to the beginning of the survey, and for the five months of the survey, the four general hospitals and the State mental hospital in the area prepared abstracts of the records of patients living in San Jose.[4] These abstracts included the name, age, sex, address, admission date, discharge date, days in hospital, surgery performed, and admission and final diagnoses. The hospitals permitted checking of their records by members of the staff of the Project to insure a complete file of abstracts. These abstracts were filed by Soundex code of the surname, and the final diagnoses were coded according to the *International Statistical Classification of Diseases, Injuries and Causes of Death.* [9]

Data on Hospitalized Illness Obtained in Household Survey

In the initial interview, household respondents were asked whether any member of the household had been a patient in a hospital overnight or longer during the 12 months preceding the month of interview. If so, data were obtained on name of hospital, month of admission, length of stay in nights, operations performed, and diagnosis.

Control cards for all families in the survey were made from the interview schedules, showing name, age, sex, and address (previous addresses were included where given). These cards, which did not contain any illness data, were filed according to the Soundex code of the surname.

Matching of Reports of Hospitalization in Survey and Hospital Records

For the initial matching operation, it was desired to locate all hospital records for all individuals included in the survey. This was done by two clerks who went through the file of control cards, systematically searching the file of hospital abstracts for persons with matching or similar names, ages, and addresses. Differences of two years or less in age were considered matching, and variations of one letter in name were disregarded. After the first check, a recheck discovered only three more matching cases.

Hospital records for "matched" persons were compared with the information obtained in the household survey. In a large proportion of

---

[4] Nearly 25 per cent of the hospitalizations of residents of San Jose occurred in hospitals outside the area. Almost half of these were by persons who reported residing in another city at some time during the year preceding the survey. The remainder were in hospitals in nearby cities, including Veterans' Administration facilities, the Southern Pacific Hospital, University of California Hospital, etc. The hospitalizations outside the area may be different in duration and type from those cases in the immediate area, so that rates given here do not represent measures of hospitalization for the population.

cases, the period of hospitalization was clearly the same in both sources, in spite of certain discrepancies in reported date of admission, length of stay, or diagnosis. In some cases, however, there were hospital records of illness which had not been reported in the household survey (these we call possible "under-reports"); other cases were reported in the household survey but not located in our file of hospital record abstracts (these we call possible "over-reports"). That is, under- and over-reporting of survey information are with respect to the criterion.

Possible over- and under-reports were subjected to further search. Variations in name were considered, and the telephone directory sometimes verified the fact that two individuals (or families) of the same name lived at different addresses in San Jose. For some cases, the name of the nearest relative was secured from the hospital to serve as a further means of identification. In this process, five additional matched cases were discovered.

Of some interest may be the degree to which a match was secured on the items used for check purposes. Table 1 summarizes the results of

TABLE 1

NUMBER OF MATCHED HOUSEHOLD SURVEY REPORTS AND HOSPITAL RECORDS AND NUMBER OF HOUSEHOLD SURVEY UNDER-REPORTS BY MATCHED ITEMS

Matched Items	Matched Records	Under-Reports
TOTAL	249	39
Name, age, address	221	33
Name, age	20	2
Name, address*	6	2
Age, address	2	2

See text for definition of terms.
* Includes those in which age was not stated in the survey.

the matching operation with respect to records of hospitalization in the five hospitals subsequent to July 1, 1951.

Under-Reporting and Over-Reporting of the Event of Hospitalization in the Household Survey

A total of 279 periods of hospitalization were reported in the household survey in the five hospitals with discharge date after July 1, 1951.[5]

[5] Persons in the city segments of the household survey reported 493 hospitalizations in the year preceding the month of interview. Of the 374 in the five hospitals in the area, 95, or 25 per cent, were

Of these, 249, or 89 per cent, were matched with hospital records. Of the 30 (11 per cent) "over-reports," i.e., reports in the household survey which were not matched with hospital records, 20 were identifiable at the hospital which was named. In eleven of these cases, the individuals had reported as during the survey period a hospitalization which actually occurred as long as one year earlier, and in seven cases, hospitalization had been reported in the household interview as overnight or longer when the hospital record showed discharge on the same day as admission. Table 2 shows a summary of over- and under-reporting.

TABLE 2

REPORTS OF HOSPITALIZATION IN HOUSEHOLD SURVEY AND IN HOSPITAL RECORDS FOR RESIDENTS OF SAN JOSE IN SAMPLE, JULY 1, 1951–MAY 31, 1952* WITH NUMBER AND PER CENT MATCHED AND WITH CLASSIFICATION OF THOSE NOT MATCHED

Category of Report	Number	Per cent
Reports of hospitalization in household survey	279	100.0
Matched with hospital records	249	89.2
Not matched with hospital records (over-reports)	*30*	*10.8*
Identified at hospital	20	7.2
Stay prior to check period	11	3.9
Stay not overnight	7	2.5
Record of this person, but not of this hospitalization	2	.7
Not identified at hospital	10	3.6
Hospitalizations in hospital records for survey population	288	100.0
Matched with household survey reports	249	86.5
Not matched with survey reports (under-reports)	*39*	*13.5*
Multiple admissions, not all admissions reported in survey	16	5.6
Reported in survey, but not during check period	8	2.8
Admissions to state mental hospital	4	1.4
Other types of under-reports	11	3.8

* Check period varied from 7 to 11 months for the different subgroups in the sample.

A total of 288 periods of hospitalization were shown in hospital records for persons in the household survey in the period subject to check.

for discharges prior to July 1, 1951, the beginning point of the time when hospital discharge records were available. The periods of time covered by the reports subject to check varied from 7 to 11 months, depending on the month in which the initial interview was taken. While episodes of hospitalization in the recent past were reported more accurately than those which occurred nearly a year before, differences in the range above 7 months were not significant. It is believed that no appreciable error was introduced by the use of the varying period subject to check.

Of these, 39, or 14 per cent, were not reported by the respondents in the household survey. In 40 per cent of these cases, persons in the survey had more than one admission in the period, and reported at least one other.[6] In another 20 per cent, the period of hospitalization was reported, but the date given was such that it did not fall within the period subject to check. There was failure to report four stays in the State mental hospital. (Two such hospitalizations were reported and matched among the 249 appearing in both survey and records.)

Comparison of Common Measures of Hospitalization as Derived from Household Sample Survey and from Hospital Records

The 279 periods of hospitalization reported in the household survey and the 288 periods disclosed in hospital records for the same population and period of time form the basis for the computation of a number of common measures of hospital utilization shown in Table 3.

TABLE 3

MEASURES OF HOSPITALIZATION FROM SAN JOSE HOUSEHOLD SURVEY AND FROM HOSPITAL RECORDS

Measure	From Survey Reports (A)	From Hospital Records (B)	Ratio B/A
Admissions per 1000 persons per year*	65.5	67.9	1.04
Days of hospitalization per person per year	.609	.655	1.08
Average length of stay per period of hospitalization in days	9.1	9.5	1.04
Per cent of admissions with surgery	43.4	44.4	1.02

Note: Because the hospitalizations upon which these rates were based were in five hospitals only, these should not be considered to represent true rates of hospitalization for the population covered.

$$^{*}\ \frac{\Sigma \text{ Admissions in period covered}}{\Sigma \text{ Person-months covered by survey}} \times 12{,}000.$$

When respondents in a household sample survey were asked about periods of hospitalization in the year preceding the survey, the information which they gave yielded an admission rate of 66 per 1000 persons in five specified hospitals. When the records of these five hospitals were searched for the names of the persons in the household survey, records were disclosed which yielded an admission rate of 68 per 1000 persons. The difference between these two rates is not significant at the

[6] In all but two of these the unreported period of hospitalization was for the same condition as the period which was reported.

five per cent level. (In all tests of significance in this paper, the five per cent level was used.)

Similarly, the days of hospitalization per person per year, the average length of stay, and the per cent of cases with surgery, were all slightly higher when obtained from hospital records than when obtained from the household survey. None of these differences was statistically significant.[7]

The difference in the average length of stay, 9.1 from survey data and 9.5 from hospital records, was accounted for by several long periods of hospitalization which were not reported in the survey. When the 239 cases in which length of stay was reported in both sources are compared, the averages become 9.2 in the survey and 8.6 from hospital records.[8] There was some indication in the data that persons in the survey tended to over-report longer stays to a greater extent than shorter stays. However, when the differences between reports from household survey and hospital records for stays of 15 days or less were compared with the differences for stays of 16 days or longer, using a *t*-test which takes into account the differences in variance at the two ends of the scale, this tendency proved to be not significant at the five per cent level [3].

Another way in which we may test the usefulness of household survey data in the reporting of hospitalization is by comparing the distributions of some of the items as reported in the household survey with the distributions obtained by reference to hospital records. Such comparisons have been made for the month of admission to hospital, length of stay, surgical procedure, and diagnosis.

Table 4 presents the month of admission of the periods of hospitalization as reported by the household survey and as obtained by the check of hospital records. Using the chi-square test, the difference between these distributions was not significant.[9]

Of some concern to medical care plans and hospital administrators is the distribution of cases by length of stay and the proportion of total days which are accounted for by stays of various lengths. The comparison of the household survey reports and hospital records as to

---

[7] Since some stays were over 200 days, the standard deviations of the distributions of stays were considerably greater than the means.

[8] It is to be noted that the household survey questions used should yield information differing from information obtained from hospitals regarding length of stay. The "number of nights" reported in the survey, if accurately reported, will always be equal to or one day less than the number of days of hospitalization shown in hospital records. The median length of stay was 4 in both distributions.

[9] There was agreement on the reported month of admission in 80 per cent of the matched cases. Discrepancies of one month appeared in fifteen per cent of the cases, and of more than one month in the remaining five per cent.

length of stay is shown in Table 5. The distribution of cases shown in the first column of Table 5 is not significantly different from the distribution shown in the second column.[10] The degree of precision which is desired in the reporting of total days of hospitalization would depend, of course, upon the uses to which the data are to be put. For

TABLE 4

279 REPORTS OF HOSPITALIZATION FROM THE SAN JOSE HOUSEHOLD SAMPLE SURVEY AND 288 FOR THE SAME POPULATION FROM HOSPITAL RECORDS, BY MONTH OF ADMISSION

Month of Admission	Survey Reports	Hospital Records
TOTAL	279	288
Prior to July, 1951	12	11
July	31	34
August	33	38
September	30	35
October	23	22
November	35	34
December	29	26
January, 1952	31	34
February	19	16
March	19	17
April	7	7
May	7	7
Not specified or not available	3	7

most practical purposes, however, it would seem that the distribution of days reported by the two methods are equally useful. It is to be noted, for example, that in the household survey 71 per cent of the total days of hospitalization were reported in stays of 60 days or less, while the corresponding figure by hospital records was 68 per cent.

When individual household survey reports are compared with the corresponding hospital records, it is found that among 239 records for which the item of length of stay was complete in both sources, there was exact agreement in 127 cases, or 53 per cent. In 65 cases, or 27 per cent, the survey report was greater than the hospital record, and

[10] When grouped into 16 categories, chi-square = 12.

in 47 cases, or 20 per cent, the survey report was less than the hospital record. The difference between the number of cases reported in the survey as more than and as less than the hospital record is not significant at the five per cent level, using the sign test [4]. It is to be remembered, however, that the definitions of length of stay differed,

TABLE 5

DISTRIBUTIONS OF PERIODS OF HOSPITALIZATION AND DAYS OF HOSPITALIZATION AS SHOWN IN THE SAN JOSE HOUSEHOLD SURVEY AND IN HOSPITAL RECORDS, BY LENGTH OF STAY IN DAYS

Length of Stay in Days*	Number of Periods		Number of Days*		Cumulative Percentages			
					Periods		Days*	
	Survey	Hospital	Survey	Hospital	Survey	Hospital	Survey	Hospital
TOTAL	279	288	2484	2672				
1	34	30	34	30	12.5	10.7	1.4	1.1
2	31	46	62	92	23.9	27.0	3.9	4.6
3	40	32	120	96	38.6	38.4	8.7	8.2
4	40	37	160	148	53.3	51.6	15.1	13.7
5	29	31	145	155	64.0	62.6	21.0	19.5
6	13	14	78	84	68.8	67.6	24.1	22.6
7	11	14	77	98	72.8	72.6	27.2	26.3
8	8	3	64	24	75.7	73.7	29.8	27.2
9	9	16	81	144	79.0	79.4	33.1	32.6
10	8	10	80	100	82.0	82.9	36.3	36.3
11	4	6	44	66	83.5	85.1	38.0	38.8
12	4	4	48	48	84.9	86.5	40.0	40.6
13	2	4	26	52	85.7	87.9	41.0	42.6
14	5	2	70	28	87.5	88.6	43.8	43.6
15–21	13	12	233	214	92.3	92.9	53.2	51.6
22–30	10	7	263	159	96.0	95.4	63.8	57.6
31–60	5	7	180	277	97.8	97.9	71.1	67.9
61 & over	6	6	719	857	100.0	100.0	100.0	100.0
Not stated	7	7	—	—				

* In the household survey reference was made to "nights in hospital."

and that a report in the survey might be accurately reported as one day less than the hospital record. If half of the household survey reports which were only one day less than the corresponding hospital record are considered as matching, the data show a tendency, which is statistically significant at the five per cent level using the sign test, for over-statement in the household survey.[11]

One item of information gathered in the survey concerned the opera-

[11] With this assumption, cases in which there was agreement would total 144, and cases in which the survey report was less than the hospital record would total 30.

tion(s) performed during a stay in the hospital. These items were coded according to the first digit of the operation code in the *Standard Nomenclature* [15]. Table 6 gives a summary of the reports from the two sources.

TABLE 6

DISTRIBUTION OF SURGICAL PROCEDURES IN 279 REPORTS OF HOSPITALIZATION FROM THE SAN JOSE HOUSEHOLD SURVEY AND 288 FROM THE HOSPITAL RECORDS

Surgical Procedure	Survey Reports	Hospital Records
TOTAL	279	288
Surgery not stated or record not available	0	9
Without surgery	158	155
With surgery	121	124*
Incision	6	7
Excision	79	92
Amputation	2	1
Introduction	1	0
Endoscopy	1	1
Repair	13	18
Destruction	3	1
Suture	1	1
Manipulation	1	3
Not classifiable above	14	0

* Twelve hospital records showed two procedures and one showed three. In these cases the procedure which matched the household survey report was counted.

In coding these procedures such terms as "sinus operation," "operated anorectal," "kidney operation," and "general surgical work" appeared in more than ten per cent of the household survey reports. These terms were not classifiable in the system used above. However, descriptions which probably referred to only one procedure, such as "hernia operation" (repair), were given the appropriate code.

It is apparent from inspection of Table 6 that the large group in the "Not classifiable" category for the household survey reports makes the picture quite different from that shown by the hospital records. This difference is statistically significant.[12] Apparently, then, household sur-

[12] The distributions in Table 6 were grouped into four categories, Incision, Excision, Repair, and Other, and the chi-square test was applied.

Note on Matched Cases: There were 240 cases in which a report of surgery was available in both

vey reports do not give as specific descriptions of surgical procedures as can be secured from hospital records.

Diagnoses given in the household survey and those on the hospital records were coded according to the 3-digit codes of the *International Statistical Classification of Diseases, Injuries and Causes of Death.* In both types of records, the diagnoses were coded in order of mention except that injuries were given precedence over other conditions. In five per cent of the survey reports there was more than one diagnosis, while there were multiple diagnoses in nearly nineteen per cent of the hospital records.

When the primary diagnoses are distributed according to a 100-group category[13] system, there are, of course, many categories in which there are few cases. These distributions are shown in Table 7. In order to test the significance of the difference here, the frequencies were grouped into 24 categories such that in no category were there less than five cases in the hospital records. Differences between the two distributions were not significant. A further consolidation of the distributions into eleven groups increased the degree of correspondence.[14]

It would appear that for most practical purposes, a description of the diagnoses of hospitalized illness to be obtained from a household survey like the one conducted in San Jose will be as useful as one obtained through reference to hospital records.

SUMMARY

In summary, most of the medical record sources which are available for validation checks on household survey reports of illness do not provide an opportunity for a comprehensive check on both the over- and under-reporting of illness. Only with respect to hospitalized illness was it possible to study the net error in reporting.

In this study, reports of hospitalization during a preceding period ranging from 7 to 11 months obtained by interview of households in

---

sources. In 133 of these (which included episiotomies and stitches incident to normal deliveries) there was a report of no surgery on both records. On one record the survey report was "Broke right ankle. Surgery: None"; while the hospital record showed "Fracture of the int. and ext. trimalleolar of right ankle, Surgery: Reduction of fractures." Here, obviously the respondent's concept of what constitutes surgery differed from that generally understood. In two other cases in which the survey reported surgery while the hospital record showed none, it appears that the hospital record may have been in error. (A) Survey report: "Osteomyelitis in left arm, operated on left arm for drainage." Hospital record: "Osteomyelitis left radius, Surgery: No." (B) Survey report: "Adhesions. Female trouble due to hysterectomy. Adhesions removed." Hospital record: "Adhesions cecum and ascending colon. Rt. Ovarian cyst with multiple varicosities. Adhesions bands intestine two to lower pelvic area, Surgery: No."

In 104 cases there was a report of surgery in both sources. In 88 of these, or eighty-five per cent, there was agreement as to the surgical procedure when classified into ten categories according to the *Standard Nomenclature.*

[13] See footnote on Table 7.

[14] In the 242 cases for which diagnoses were available both from the household survey and from the hospital record, there was agreement between one pair of diagnoses at the 800-group level in 61 per cent of the cases. At the 100-group level, this agreement increased to 76 per cent, and when the data are arranged into 15 groups, the agreement was 85 per cent.

San Jose, California, were checked against hospital records in five hospitals. Records were matched for 249 periods of hospitalization. Thirty reports in the survey could not be matched with hospital records. These over-reports included hospitalization which was not overnight, stays which occurred as long as one year earlier than the study period, as well as ten which could not be identified at the named hospital. Thirty-nine periods of hospitalization were not reported by the respondents in the survey. Nearly half of these under-reports were for persons with multiple admissions, who reported at least one other hospitalization. Four were unreported admissions to the State mental hospital.

Admission rates based on survey reports did not differ significantly from those based on hospital records for the same population. Similarly, days of hospitalization per person per year, average length of stay per period of hospitalization, and per cent of admissions with surgery were calculated accurately from household survey data.

Distributions of admissions from household survey reports of hos-

TABLE 7

DISTRIBUTION OF SOLE OR PRIMARY DIAGNOSES IN 279 REPORTS OF HOSPITALIZATION FROM THE SAN JOSE HOUSEHOLD SURVEY AND 281* FROM THE HOSPITAL RECORDS

ISC[1] Codes	Diagnostic Category[2]	Number of Hospitalizations		Per cent of Total	
		Survey Reports	Hospital Records	Survey Reports	Hospital Records
	TOTAL	279	281*	100	100
001–138	*Infective and parasitic diseases*	*10*	*7*	3.6	2.5
	Tuberculosis of respiratory system	2	1		
	Food poisoning	1	0		
	Acute poliomyelitis, infectious encephalitis and late effects	4	4		
	Other infective and parasitic diseases	3	2		
300–326, 353, 780, 781, 790, 791	*Psychoneuroses, mental disorder and ill-defined nervous conditions*	*6*	*10*	2.2	3.6
	Mental, psychoneurotic and personality disorders	6	9		
	Epilepsy	0	1		
370–389	*Diseases of eyes*	*2*	*3*	.7	1.1
	Other inflammation of eye	0	1		
	Cataract	2	1		
	Other diseases of eye	0	1		
363, 400–416, 720–727	*Rheumatic fever, arthritis, muscular rheumatism and sciatica*	*3*	*2*	1.1	.7
	Rheumatic fever and chronic rheumatic heart disease	1	0		
	Arthritis, not elsewhere classified	2	2		

TABLE 7—(*continued*)

ISC[1] Codes	Diagnostic Category[2]	Number of Hospitalizations		Per cent of Total	
		Survey Reports	Hospital Records	Survey Reports	Hospital Records
420–468, 782	*Diseases of circulation and symptoms referable to it*	*24*	*22*	8.6	7.8
	Arteriosclerotic heart and coronary disease	4	7		
	Other diseases of heart	6	2		
	Hypertensive disease	4	1		
	Diseases of arteries	3	2		
	Varicose veins of lower extremities	2	2		
	Haemorrhoids	4	6		
	Other diseases of circulatory system	0	1		
	Symptoms referable to cardiovascular and lymphatic system	1	1		
470–483	*Colds, influenza and acute respiratory infections*	*1*	*2*	.4	.7
	Other acute upper-respiratory infections	0	1		
	Influenza	1	1		
490–527, 783	*Other respiratory diseases and symptoms*	*21*	*17*	7.5	6.0
	Pneumonia	6	5		
	Hypertrophy of tonsils and adenoids	13	8		
	Chronic sinusitis and nasal diseases	1	2		
	Pleurisy, empyema and lung abscess	0	1		
	Symptoms referable to respiratory system	1	1		
540–545, 784	*Disorders of upper gastro-intestinal tract*	*8*	*8*	2.9	2.8
	Ulcer of stomach	5	1		
	Ulcer of duodenum	0	5		
	Other disorders of stomach and duodenum	2	2		
	Symptoms referable to upper gastro-intestinal tract	1	0		
550–587, 785	*Disorders of lower gastro-intestinal tract*	*30*	*33*	10.8	11.7
	Appendicitis	12	10		
	Hernia and intestinal obstruction	7	8		
	Gastro-enteritis and colitis	2	7		
	Cholelithiasis and cholecystitis	3	4		
	Other disorders of digestive system	6	4		
590–689, 786	*Disorders of genito-urinary system and complications of childbearing*	*97*	*97*	34.8	34.5
	Nephritis	1	0		
	Calculi of urinary system	1	0		
	Other diseases of urinary system	3	5		
	Hyperplasia of prostate	0	1		
	Other diseases of male genital organs	2	0		
	Diseases of breast	2	0		
	Disorders of menstruation and menopause	0	2		
	Other diseases of female genital organs	11	10		
	Complications of pregnancy, childbirth and puerperium	77	78		
	Symptoms referable to genito-urinary system	0	1		

TABLE 7—(*continued*)

ISC[1] Codes	Diagnostic Category[2]	Number of Hospitalizations		Per cent of Total	
		Survey Reports	Hospital Records	Survey Reports	Hospital Records
690–716, 730–759	*Diseases of skin and skeletal system, malformations*	*10*	*14*	3.6	5.0
	Other infections of skin and subcutaneous tissue	0	2		
	Other diseases of skin and subcutaneous tissue	1	1		
	Osteomyelitis and periostitis	1	1		
	Other diseases of bone	2	1		
	Other diseases of joints except ankylosis	5	2		
	Other acquired musculoskeletal deformities	1	3		
	Congenital malformations	0	4		
140–299, 330–334, 340–352, 354–362, 364–369, 760–776, 787–789, 792–795	*Other diseases and symptoms*	*47*	*49*	16.8	17.4
	Neoplasms	30	28		
	Asthma	1	0		
	Diabetes mellitus	3	3		
	Anaemias	0	1		
	Other allergic, endocrine, metabolic and blood diseases	1	3		
	Vascular diseases affecting central nervous system	3	4		
	Other diseases of central nervous system	1	3		
	Diseases peculiar to early infancy	2	2		
	Symptoms referable to limbs and back	0	1		
	Other and ill-defined symptoms and conditions	6	4		
N800–999	*Injuries*	*20*	*17*	7.2	6.0
	Fractures	7	8		
	Dislocations, sprains and strains	3	0		
	Interval injury of chest, abdomen, pelvis and head injury without fracture	1	1		
	Lacerations and open wounds	1	1		
	Burns	1	1		
	Other and unspecified effects of external causes	7	6		

[1] See reference [9]. Groups used are from "Drafts of Five Special Condensations and Expansions of the International Statistical Classification of Diseases, Injuries, and Causes of Death to Provide for Presentation in Convenient Form of Statistics of Sickness Surveys, Sickness Absenteeism, and Hospital Diagnosis," received from I. M. Moriyama, Secretary, U. S. Committee of Vital and Health Statistics, with letter dated 1-6-53.

[2] Categories not included in which there were no reported periods of hospitalization.

[3] Excludes 7 cases for which diagnoses were not available.

pitalizations by month of admission, length of stay, and diagnosis were similar to the distributions obtained from hospital records. Whether or not surgery was performed was reported accurately in the household survey, but the description of surgical procedure was not as precise as that obtained from hospital records.

This study shows that reports of hospitalization obtained in household sample surveys are sufficiently accurate to be used for many purposes in lieu of hospital record data.

REFERENCES

[1] Breslow, Lester, "California morbidity research project—Conduct of the San Jose survey," to be published.
[2] Collins, Selwyn D., "Causes of illness in 9,000 families based on nation-wide periodic canvasses, 1928–1931," *Public Health Reports*, 48 (1933), 283–308.
[3] Dixon, Wilfrid J., and Massey, Frank J., Jr., *Introduction to Statistical Analysis*, New York, McGraw-Hill, (1951), pp. 104–5.
[4] ———, pp. 247–252.
[5] Downes, Jean, and Collins, Selwyn D., "A study of illness among families in the Eastern Health District of Baltimore," *Milbank Memorial Fund Quarterly*, 18 (1940), 5–26.
[6] Downes, Jean, Collins, Selwyn D., and Jackson, E. H., "Medical care among males and females at specific ages—Eastern Health District of Baltimore, 1938–1943," *Milbank Memorial Fund Quarterly*, 29 (1951), 5–30.
[7] Falk, I. S., Klem, Margaret C., and Sinai, N., *The Incidence of Illness and the Receipt and Costs of Medical Care Among Representative Families*, Chicago, University of Chicago Press, (1933), p. 72.
[8] Holland, Dorothy F., and Altenderfer, Marion E., *Sickness in a Metropolitan Community—The Results of a Health Survey of New York City*, U. S. Public Health Service, multilithed, 161 pp.
[9] *International Statistical Classification of Diseases, Injuries and Causes of Death*, Sixth Revision of the International List of Diseases and Causes of Death, Geneva, World Health Organization, (1949), Two vols.
[10] Jackson, Elizabeth H., "Duration of disabling acute illness among employed males and females—Eastern Health District of Baltimore, 1938–1943," *Milbank Memorial Fund Quarterly*, 29 (1951), 294–330.
[11] Kennedy, M. Eileen, "Administration and methods of enumeration of the sickness survey in Alberta," *Canadian Journal of Public Health*, 44 (1953), 177–79.
[12] Marks, Eli S., and Mauldin, W. Parker, "Response errors in census research," *Journal of the American Statistical Association*, 45 (1950), 424–38.
[13] Peart, A. F. W., "Canada's sickness survey—Review of methods," *Canadian Journal of Public Health*, 43 (1952), 401–14.
[14] Perrott, George St. J., Tibbitts, Clark, and Britten, Rollo H., "The National health survey," *Public Health Reports*, 54 (1939), 1663–87.
[15] *Standard Nomenclature of Diseases and Operations*, American Medical Association, New York, Blakiston Co. (1952), p. 518.
[16] Stecker, Margaret L., *Some Recent Morbidity Data*, New York, Metropolitan Life Insurance Co. (1919), p. 12.
[17] Sydenstricker, Edgar, "A study of illness in a general population group," *Public Health Reports*, 41 (1926), 2069–88.
[18] Weissman, Arthur, "California morbidity research project," *American Journal of Public Health*, 42 (1952), 711–16.

BUSINESS FAILURES: ANOTHER EXAMPLE OF THE ANALYSIS OF FAILURE DATA

K. S. Lomax
University of Manchester

The analyses of failure data given by Davis [1] all involve essentially constant or increasing conditional probabilities of failure. For business failures, however, it is reasonable to expect monotonically decreasing conditional probabilities. An analysis of data on failures of four types of business in Poughkeepsie, New York, from 1844 to 1926 [2] confirms this expectation. The conditional probabilities of failure for these four series are well described by both exponential and hyperbolic functions.

The interesting and stimulating paper by D. J. Davis [1] on failure data is most suggestive to the economist.

Broadly, Davis analyzes three types of failure theory:

(*a*) The normal theory of failure, in which the failure probability density function is Gaussian.

(*b*) Human mortality, characterized by rapid increase of the conditional density function after middle-age.

(*c*) Exponential theory of failure, in which the conditional density function is constant.

In (*a*) uniformly and in (*b*) after the very early years of life the conditional density function of failure probability with time is strictly monotonic increasing. In (*c*) it is constant.

The economist immediately thinks of business failures in which it is reasonable to expect the conditional density function strictly to *decrease* monotonically. The purpose of this note, then, is to draw attention to a fourth category of failure theory:

(*d*) Business mortality. It is fairly well established that with most types of business the early years are the most difficult. It is then that mortality is highest. The longer a business survives, generally, other things being equal, the smaller becomes the probability of failure.

Take, for example, the useful and comprehensive data compiled by R. G. and A. R. Hutchinson and Mabel Newcomer [2]. Their Table I, showing the length of life for business enterprises established in Poughkeepsie between 1844 and 1926, can serve as basis for calculation of $F(t)$, for different values of t, where

$$F(t) = \text{cumulative probability of failure in the interval } (0, t).$$

The results of these calculations, omitting wholesale businesses since the sample there was small in comparison with the other categories, are shown in Table 1.

TABLE 1

CUMULATIVE PROBABILITY OF BUSINESS FAILURE IN POUGHKEEPSIE, 1844–1926

Age in years	Retail	Manufacture	Craft	Service
1	0.296	0.231	0.307	0.327
2	0.438	0.346	0.454	0.457
3	0.532	0.469	0.551	0.551
4	0.594	0.547	0.607	0.618
5	0.643	0.602	0.660	0.669
6	0.684	0.655	0.697	0.708
7	0.715	0.678	0.727	0.743
8	0.741	0.702	0.753	0.769
9	0.762	0.726	0.772	0.792
10	0.782	0.746	0.791	0.812

Source: Calculated from Hutchinson, Hutchinson, and Newcomer [2].

In all these cases the graph of $F(t)$ takes the shape shown in Figure 1.

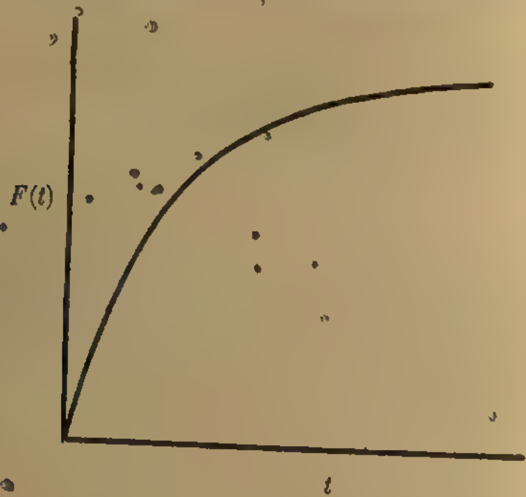


Fig. 1

Now, if

$f(t)$ = probability density function of failure time
= probability of failure in infinitesimal interval $(t, t+dt)$,

then $f(t) = F'(t)$ and two methods are available for estimation of $f(t)$ from the above data. One is to measure the slopes of the $F(t)$ graphs at the different values of t. The other is to use the difference formula

$$F'(t) = \Delta F(t) - \tfrac{1}{2}\Delta^2 F(t) + \tfrac{1}{3}\Delta^3 F(t) - \cdots .$$

The former method seems to be the more satisfactory here, and the results of applying it to the data of Table 1 are shown in Table 2.

TABLE 2

PROBABILITY DENSITY OF BUSINESS FAILURE IN POUGHKEEPSIE, 1844–1926

Age of years	Retail	Manu-facture	Craft	Service
0	0.57	0.365	0.5	0.5
1	0.175	0.152	0.213	0.192
2	0.107	0.119	0.102	0.106
3	0.071	0.090	0.072	0.081
4	0.056	0.067	0.056	0.059
5	0.045	0.051	0.044	0.044
6	0.036	0.037	0.034	0.038
7	0.028	0.030	0.028	0.031
8	0.023	0.023	0.024	0.025
9	0.020	0.014	0.020	0.021
10	0.018	0.007	0.017	0.019

Source: Computed from Table 1.

Thus, $f(t)$ can generally be represented as in Figure 2.

FIG. 2

From the values of $f(t)$ and $F(t)$, following Davis, we calculate $Z(t)$ the conditional density function of failure probability with time, in other words, the instantaneous probability rate of failure at time t conditional upon non-failure prior to t:

$$Z(t) = \frac{f(t)}{1 - F(t)}.$$

These results are shown in Table 3.

TABLE 3
CONDITIONAL PROBABILITY DENSITY OF BUSINESS FAILURE, POUGHKEEPSIE, 1844–1926

Age in years	Retail	Manu-facture	Craft	Service
0	0.57	0.365	0.5	0.5
1	0.249	0.198	0.307	0.285
2	0.190	0.182	0.187	0.195
3	0.152	0.169	0.160	0.180
4	0.138	0.148	0.142	0.154
5	0.126	0.128	0.129	0.133
6	0.114	0.107	0.112	0.130
7	0.098	0.093	0.103	0.121
8	0.089	0.077	0.097	0.108
9	0.094	0.051	0.088	0.101
10	0.083	0.028	0.081	0.101

Source: Calculated from Tables 1 and 2.

$Z(t)$ is of the form shown in Figure 3.

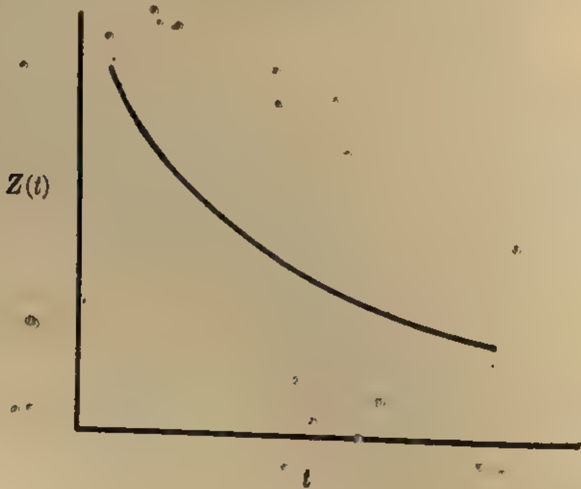


FIG. 3

There is really little purpose to be served by searching for analytical expressions representing this behavior. This could only be useful if it were feasible to obtain general support from extraneous sources for a particular form of relationship. The only such support in this case is in relation to such trivialities as

$$F(t), f(t), Z(t) \ngtr 1,$$
$$F(0) = 0; \quad f(0) = Z(0),$$
$$F(t) \text{ monotonic increasing.}$$

It is, however, of interest to record that a good fit to the $Z(t)$ values can be obtained, in each case, either by the exponential function

$$Z(t) = ae^{-bt}$$

or the hyperbola

$$Z(t) = \frac{b}{t + a}.$$

The latter appears to be the more appropriate for the Retail, Craft, and Service groups, while in the case of Manufacturing trades the exponential gives the better fit. These above functions were fitted to the data in the transformations

$$\log_e Z(t) = \log_e a - bt, \quad \text{linear in } t \text{ and } \log_e Z(t)$$

and

$$\frac{1}{Z(t)} = \frac{1}{b}t + \frac{a}{b}, \quad \text{linear in } t \text{ and } \frac{1}{Z(t)}.$$

The correlation coefficients corresponding to these linear forms are shown in Table 4.

TABLE 4

CORRELATION COEFFICIENTS FOR FUNCTIONS FITTED TO DATA OF TABLE 3

Type of business	Exponential	Hyperbola
Retail	0.91	0.99
Manufacture	0.96	0.83
Craft	0.93	0.99
Service	0.91	0.98

One advantage of the hyperbolic "law" is that the expressions for $f(t)$ and $F(t)$ remain fairly simple.

If

$$Z(t) = \frac{b}{t+a},$$

then

$$F(t) = 1 - \left(\frac{a}{t+a}\right)^b$$

and

$$f(t) = \frac{b}{a}\left(\frac{a}{t+a}\right)^{b+1};$$

whereas

$$Z(t) = ae^{-bt}$$

leads to

$$F(t) = 1 - e^{a/b(e^{-bt}-1)}$$

and

$$f(t) = ae^{-bt+a/b(e^{-bt}-1)}.$$

Both alternatives conform to the desirable boundary conditions and monotonic behavior exhibited by the data.

I hope shortly to carry out a more detailed analysis of business mortality covering British as well as American data. A cursory examination has already indicated that British experience does not always completely accord with American.

REFERENCES

[1] Davis, D. J., "An analysis of some failure data," *Journal of the American Statistical Association*, 47 (1952), 113–50.

[2] Hutchinson, Ruth Gillette and Arthur R., and Newcomer, Mabel, "Study in business mortality," *American Economic Review*, 28 (1938), 497–514.

CYCLICAL FLUCTUATIONS IN FOUNDRY ACTIVITY

LLOYD SAVILLE
Duke University

I. INTRODUCTION

THE extreme sensitivity of foundry operations to business change has been apparent for many years [7]. Only recently, however, has sufficient information been available to permit an analysis of the cyclical movements of foundry output as a whole. An adequate accumulation of data in a number of the *Facts for Industry Series* of the Bureau of the Census, some of it collected for the first time during World War II, now makes possible the construction of a seasonally adjusted index of foundry activity covering a period of business fluctuations.

In general, foundries produce metal parts or castings to the custom specification of local firms in durable goods industries; consequently foundry activity is influenced by cyclical fluctuations in a wide range of geographic and industrial areas. Five general characteristics of the industry affect its sensitiveness to change: (1) The production of foundries is dominated by changes in the demand for a variety of products commonly classified as durable consumers' goods, investment goods, and war materials: castings are employed as bases for pumps, lathes, and presses; as frames for pianos, lawnmowers, and locomotives; as wheels for railroad cars, airplanes, and machines; and as component parts for lamps, engines, and motors. (2) Castings are made by several thousand establishments operating in many geographically separated markets; under these atomistic conditions, foundry production describes the activities of a wide range of firms and tends to minimize the influence of one or a few firms on production totals. (3) The production of castings made from different metals is dominated by various technical requirements: use of aluminum castings is regulated by the level of output of such objects as airplanes and portable tools, where lightness of weight is a major consideration; the quantity of castings made from brass and bronze, malleable iron, steel, and gray iron reflects, respectively, the rate of manufacture of corrosion-resistant fittings needed on ships and in chemical plants, shock-withstanding parts for railroads, armor plates for tanks, and general components for producers' and consumers' goods. (4) Since inventories of rough castings tend to be small, the production of castings is closely related to current demand. Foundrymen usually make parts to the specification of the individual consumer and so find it difficult to accumulate castings

for future orders. (5) Finally, the makers of castings form not only a variety of items but also a sizeable product. In 1947 their value added to product totaled almost two billion dollars, or approximately 2.5 per cent of the value added by the producers of all the manufactured goods in the United States (Table 1).

TABLE 1

ESTIMATES OF ACTUAL AND RELATIVE VOLUME OF PRODUCT, VALUE OF PRODUCT, AND VALUE ADDED TO PRODUCT BY CERTAIN CLASSES OF FOUNDRIES AND ALL MANUFACTURERS IN THE UNITED STATES, 1947

Group of manufacturers	Actual figures: Quantity of product in pounds (000,000)	Actual figures: Value of product in dollars (000,000)	Actual figures: Value added to product in dollars (000,000)	Percentages: Quantity of product	Percentages: Value of product	Percentages: Value added to product
Aluminum and aluminum base foundries	468	256	135	1.7	8.4	7.3
Copper and copper base foundries	1,061	394	207	3.8	12.9	11.2
Malleable-iron foundries	1,797	231	152	6.4	7.6	8.2
Steel foundries	2,532	389	252	9.0	12.8	13.6
Gray-iron foundries	22,271	1,773	1,108	79.1	58.3	59.7
Total of all foundries	28,129	3,043	1,854	100.0	100.0	100.0
Total of all manufacturers	not available	not available	74,426	—	—	—

Source: [12, Vol. II, pp. 21, 535, 546, 560, and 564]. Data adjusted to make coverage comparable with quantities presented in Table 3.

2. AVAILABLE FOUNDRY INFORMATION

Almost all of the monthly information concerning the operations of foundries in the United States is found in the *Current Statistical Service* and the *Facts for Industry* series of the Bureau of the Census. Data for brief periods and for special groups of producers have been collected by other governmental agencies (*e.g.*, Office of Price Administration) and trade associations (*e.g.*, Malleable Founders' Society). Table 2 lists the dates and designations of publications of the Bureau of the Census in which monthly reports of foundry activity are available. Series dating from 1923 and 1926 are shown for malleable iron and steel, but only from 1942 and 1943 for castings made from aluminum, copper,

and gray iron. Consideration has not been given to castings of magnesium and lead for they are of minor importance, comprising in total about 2 per cent (by weight) of all nonferrous castings shipped in 1953. Although castings made from zinc approximate in volume castings made from aluminum, they have been excluded from the inquiry for two reasons: (1) Comparable data concerning them is not available prior to 1946. (2) And, more important, the technology and market structure associated with die castings, the predominant way of forming zinc, is quite different from the techniques associated with the sand and permanent molding methods used in making castings of other metals.

TABLE 2

BUREAU OF THE CENSUS PUBLICATIONS OF MONTHLY FOUNDRY OPERATING INFORMATION IN THE UNITED STATES, 1923 TO 1954

Metal	Start of series	Publications in which data are included (Dates are inclusive.)	Designations of current publications
Aluminum	January, 1942	Series 1-1, 1-3, 1-6, and 1-7, Jan. 1942 to Sept. 1945; Series M24B, Oct. 1945 to Dec. 1945; Series M24E, Jan. 1946 to present.	M24E
Copper	January, 1942	Series 1-1, 1-3, 1-6, and 1-7, Jan. 1942 to Dec. 1945; Series M24E, Jan. 1946 to present	M24E
Malleable Iron	May, 1923	Current Statistical Service, May, 1923 to June, 1944; Series 30-7, Jan. 1943 to Sept. 1945; Series M21B, Oct. 1945 to Dec. 1950; Series M21-1 and M21C, Jan. 1950 to present.	M21-1 and M21C
Steel	January, 1926	Current Statistical Service, Jan., 1926 to June, 1944; (Production of Commercial Steel Castings, only) Series 30-1, July, 1943 to Sept. 1945; Series M22A, Oct. 1945 to Dec. 1950; Series M21-1 and M21C, Jan. 1951 to present.	M21-1 and M21C
Gray Iron	January, 1943	Series 30-2, Jan. 1943 to June 1944; Series 30-5, July, 1944 to Sept. 1945; Series M21A, Oct. 1945 to Dec. 1950; and Series M21-1 and M21C, Jan. 1951 to present. (Miscellaneous castings series first available Oct. 1944)	M21-1 and M21C

Currently, only unfilled-order and shipment data are being published; shipment figures are the more descriptive of the two. Unfilled orders tend to fluctuate not only with the demand for castings, but also with the availability of casting facilities, for customers often place duplicate orders with a number of foundries during boom times in

the hope of achieving prompt delivery from one of them. This complexity, and the fact that figures are not available prior to December, 1945 for malleable iron and steel castings, reduces the value of unfilled orders as a measure of cyclical change. Since inventories of finished castings held by foundries usually are not large, shipment data correspond very closely to production information; when both production and shipment figures are available concurrently, no material differences are evident.

Shipment series for the five foundry industries, inflated to achieve full coverage and comparability, are shown in Table 3. The universe of firms has been ascertained with some precision, for the reports submitted by each company during World War II were used in connection with the allocation of scarce materials; even establishments which may have dealt in the black market probably complied with the filing requirements of the regulations in order to secure a legitimate quota of metal.

The coverage of monthly releases varies widely; in the case of malleable-iron foundries, virtually every producer reported every month during the last ten years, while in the case of each of the other foundry industries, reports were collected from all known producers in 1946 and 1950. Between these complete enumerations, samples of varying size were assembled each month. The Census Bureau revised the monthly reports for 1945 and 1946 on the basis of the annual reports collected from the universe of firms in 1946, and revised the monthly reports for 1948 and 1950 on the basis of the 1950 study. Similar revisions have been made in the monthly data for 1947 and 1948 by the author based on adjustments of annual totals for these years published by the Census Bureau. A new sample of the nonferrous firms, makers of aluminum and copper castings, was established in September, 1952; the monthly totals for 1951 and 1952 have been revised on this basis.

With the exception of the gray-iron industry the data reflect activity of all of the firms in each of these industries. Shipments of gray iron comprise only three of the five groups of producers often classified in the broad category, gray-iron foundries. Miscellaneous gray-iron castings, molds for heavy steel ingots, and chilled iron railroad car wheels are included in the series because the techniques involved in their manufacture and sale are generally similar to the processes of other foundries studied here; cast iron pressure pipe and fittings and cast iron soil pipe and fittings are excluded for they are standardized-inventory items which are made and marketed quite differently from the custom-made products included in the study.

TABLE 3

MONTHLY SHIPMENTS OF ALUMINUM, COPPER, MALLEABLE-IRON, STEEL, AND GRAY-IRON CASTINGS IN THE UNITED STATES, UNADJUSTED FOR SEASONAL FLUCTUATIONS

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
					ALUMINUM—Pounds (000,000)							
1942	20.04	19.82	23.41	25.00	24.14	25.56	27.09	29.26	30.79	33.94	32.53	34.24
1943	33.26	33.56	38.94	37.26	37.90	36.50	37.36	38.38	41.06	43.01	43.18	41.61
1944	44.40	45.90	49.40	42.90	44.60	42.40	38.60	43.20	41.40	42.10	40.50	39.00
1945	45.59	44.09	50.28	46.32	43.32	39.07	31.13	22.84	15.80	19.82	18.96	17.77
1946	24.58	23.58	27.94	29.70	31.21	30.86	31.16	36.38	34.56	43.01	38.23	37.50
1947	43 07	39.98	44.74	44.11	39.59	34.92	31.97	32.22	37.98	42.90	36.71	39.68
1948	42.10	42.17	46.43	42.52	37.63	39.80	32.15	35.71	39.86	39.49	38.39	35.37
1949	33.28	31.18	31.54	27.45	24.62	26.77	21.44	27.72	31.91	33.83	30.79	31.25
1950	33.81	34.01	42.58	39.77	42.98	44.86	35.95	47.83	50.69	57.45	56.88	56.28
1951	50.59	48.52	50.18	47.84	48.04	42.58	32.44	41.76	38.16	41.31	37.84	35.88
1952	40.22	40.03	41.44	43.97	42.09	39.87	35.28	39.20	45.43	51.63	46.48	53.34
1953	55.63	54.86	59.42	61.58	57.60	56.82	51.65	50.90	53.69	55.44	51.40	52.02
1954	50.93	51.21	56.18	53.01								
					COPPER—Pounds (000,000)							
1942	113.60	110.60	113.80	115.00	108.00	108.60	112.80	108.00	123.80	127.60	110.20	113.80
1943	107.60	114.20	125.60	127.60	117.40	143.00	123.20	127.20	132.80	135.60	133.00	134.80
1944	134.40	138.80	148.60	134.80	136.40	134.20	119.06	128.47	122.82	124.94	121.41	109.18
1945	123.06	116.94	132.24	110.76	116.00	102.35	96.00	89.18	68.24	80.24	76.24	65.65
1946	84.56	74.85	85.24	91.36	89.16	83.16	79.38	93.30	92.37	106.48	92.71	93.03
1947	102.76	93.46	95.44	99.20	90.92	85.41	71.28	75.78	83.96	93.01	82.59	88.06
1948	87.06	88.83	98.17	90.74	82.84	86.44	73.17	83.54	88.63	93.05	88.15	89.86
1949	77.61	70.09	71.31	60.69	54.66	54.00	44.47	59.74	62.43	59.85	63.48	65.50
1950	67.33	68.60	82.66	74.41	88.47	85.71	71.45	97.45	99.25	109.71	107.49	104.43
1951	106.56	102.05	116 77	107.44	111.24	104.18	81.26	100.61	91.06	101.07	93.64	81.55
1952	90.85	87.58	90.22	85.96	83.40	76.77	66.18	78.80	85.82	93.10	80.44	90.80
1953	85.78	85.18	92.80	94.36	85.22	84.66	71.04	78.36	80.31	83.66	74.71	77.67
1954	71.44	68.85	76.48	72.90								
					MALLEABLE IRON—Short Tons (000)							
1923	n.a.	n.a.	n.a.	n.a.	80.41	80.79	75.34	81.81	63.78	76.57	63.62	59.29
1924	82.60	75.12	79.03	69.88	66.57	47.27	45.02	43.41	47.11	52.84	51.25	59.99
1925	68.41	69.65	76.12	75.73	74.54	72.35	66.90	64.41	66.66	73.88	68.47	72.00
1926	65.03	73.38	81.42	77.32	68.83	72.84	64.74	66.74	64.22	63.07	53.05	53.09
1927	58.54	64.98	76.56	71.43	67.76	66.63	57.99	61.15	56.37	52.09	48.09	53.45
1928	59.26	65.23	76.15	69.50	70.37	71.36	64.24	71.59	66.01	67.90	62.39	60.80
1929	82.57	76.01	86.67	86.57	86.10	77.35	74.44	74.64	66.90	62.79	49.71	50.99

TABLE 3—(continued)

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1930	61.82	64.12	69.67	65.09	59.[illegible]	46.98	34.92	34.04	31.24	31.23	27.77	29.37
1931	32.70	34.40	38.85	40.58	38.95	30.52	25.66	22.37	20.09	19.11	19.09	20.93
1932	22.34	22.08	22.03	19.15	20.55	18.69	12.91	9.81	11.08	13.20	14.48	16.33
1933	16.27	16.16	12.50	19.61	26.23	33.26	33.14	34.34	28.86	23.20	22.36	25.35
1934	30.27	35.69	47.19	45.25	44.57	35.92	31.35	29.30	23.14	24.64	24.57	33.62
1935	46.80	42.78	48.84	52.37	42.[illegible]	36.25	35.35	31.56	38.00	45.60	47.08	49.93
1936	49.31	44.75	49.80	58.91	55.51	52.83	52.45	40.40	45.67	54.93	56.86	65.47
1937	58.81	63.34	76.43	71.48	65.15	64.68	50.82	49.77	54.25	49.72	42.08	31.45
1938	23.76	23.41	26.09	26.18	23.36	24.62	19.89	23.98	30.61	32.64	40.41	41.40
1939	39.14	37.31	42.81	36.20	35.12	35.02	28.14	35.80	42.17	53.56	58.11	57.80
1940	56.01	47.25	46.22	45.14	44.00	35.83	36.81	46.47	49.40	65.7[illegible]	60.56	64.66
1941	70.84	66.74	72.49	78.57	77.14	75.46	73.45	69.09	72.61	88.17	74.17	76.06
1942	70.58	67.44	70.83	73.61	66.43	63.59	63.57	60.91	63.42	70.39	62.88	68.49
1943	59.56	67.90	76.53	70.74	69.15	70.58	67.95	68.48	71.87	72.21	72.84	76.83
1944	74.45	73.23	81.22	69.36	72.28	71.76	61.70	70.17	72.82	76.88	77.53	76.83
1945	78.79	75.22	85.31	76.06	79.56	71.99	55.81	52.65	46.96	59.10	57.32	51.96
1946	54.02	40.16	50.24	65.01	62.60	61.65	64.45	67.90	69.51	79.21	68.99	68.31
1947	76.22	75.00	78.92	82.22	75.80	78.85	64.43	62.66	71.57	84.33	72.41	78.08
1948	78.59	75.86	87.50	81.29	76.72	82.44	65.55	73.89	78.48	82.45	77.85	80.56
1949	72.69	67.53	72.92	62.09	55.29	60.36	44.96	58.95	61.61	58.02	50.20	58.28
1950	64.11	61.57	67.83	71.25	77.91	84.31	69.09	88.20	84.54	92.20	87.36	93.84
1951	92.51	88.95	102.17	97.92	101.34	94.38	76.83	90.73	82.28	93.88	88.21	76.04
1952	87.00	82.90	80.96	89.27	81.77	74.45	45.27	63.72	75.95	88.06	76.10	80.68
1953	87.25	86.52	94.48	95.92	82.05	86.51	77.11	73.86	74.33	73.47	63.44	72.13
1954	70.29	69.08	84.34	74.52								
					STEEL—Short Tons (000)							
1926	156.3	154.5	186.3	183.4	166.3	159.8	146.6	137.3	133.7	134.0	143.2	137.3
1927	141.0	143.9	166.3	153.2	140.6	142.1	129.0	141.0	113.5	101.3	95.8	94.7
1928	120.2	141.5	151.6	138.7	150.5	147.9	126.8	141.5	122.3	141.9	132.9	132.4
1929	150.6	157.4	185.8	196.3	204.8	187.1	190.0	195.0	172.1	194.5	177.4	171.5
1930	175.8	173.5	183.1	178.5	169.4	146.5	127.4	103.7	99.8	96.0	71.5	74.7
1931	74.7	79.8	91.6	77.9	69.7	56.5	51.3	48.7	43.4	38.9	37.3	34.2
1932	29.5	29.7	32.6	27.6	22.9	20.2	18.5	19.8	18.2	20.2	22.1	22.1
1933	22.6	20.3	21.9	19.9	30.6	43.4	46.6	49.7	41.3	40.9	36.3	34.9
1934	40.6	41.9	58.1	67.9	84.3	74.0	67.9	64.3	46.8	42.8	37.9	35.1
1935	40.3	41.2	44.3	44.4	42.5	38.5	43.2	84.6	49.2	59.2	50.3	52.5
1936	61.5	66.7	71.8	87.6	89.2	97.6	109.3	113.3	106.4	103.9	95.7	116.1
1937	118.2	124.5	147.0	139.1	126.3	133.2	114.5	121.2	109.2	86.8	67.5	54.6
1938	40.8	36.1	40.5	33.2	29.1	31.7	30.3	48.0	37.5	37.0	39.5	47.5
1939	52.4	49.5	54.1	47.6	54.5	53.0	45.0	55.8	57.4	94.9	112.9	104.9

TABLE 3—*(continued)*

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1940	105.4	88.8	76.8	68.6	65.8	66.7	76.1	87.4	84.3	110.4	106.8	112.9
1941	124.2	112.5	125.3	134.2	138.2	150.0	147.9	154.9	155.9	178.0	137.6	173.0
1942	177.4	175.9	192.8	196.4	173.0	173.8	178.6	183.2	183.9	200.1	184.7	189.3
1943	208.9	208.9	208.8	215.7	208.9	208.6	208.0	193.5	212.6	209.3	213.2	224.8
1944	250.1	223.2	236.7	215.0	221.5	210.7	186.0	215.6	201.5	210.4	209.0	203.0
1945	210.2	191.4	222.6	197.7	192.9	173.7	139.3	131.4	114.6	130.3	123.0	115.2
1946	102.5	60.6	104.5	146.6	130.8	121.5	117.5	129.7	126.5	137.3	130.6	123.9
1947	139.1	125.7	135.0	144.3	141.0	139.1	117.0	120.4	137.6	148.5	130.2	148.2
1948	141.2	142.5	163.1	150.5	143.4	153.1	120.5	140.3	149.3	153.2	146.9	157.6
1949	141.1	138.4	142.9	116.7	109.0	118.7	79.0	90.3	87.5	71.6	76.8	84.9
1950	88.8	92.4	112.7	108.0	119.3	133.6	99.1	130.3	136.4	152.4	149.0	158.7
1951	166.5	157.0	182.8	175.8	180.2	177.9	140.3	177.1	160.7	189.9	176.7	165.1
1952	139.5	183.7	173.7	175.1	173.6	141.6	119.0	150.2	158.4	165.2	148.3	161.7
1953	167.2	175.7	182.2	179.6	165.6	164.7	139.6	141.3	135.3	140.7	114.1	123.3
1954	122.8	116.5	122.3	105.8								
GRAY IRON—Short Tons (000)												
1943	669.8	659.4	774.5	817.8	803.1	817.7	710.3	743.6	798.3	798.0	765.7	799.8
1944	776.3	773.9	837.0	764.6	788.3	760.1	681.7	770.1	736.6	767.1	751.8	736.6
1945	804.4	761.6	858.9	779.3	800.5	778.8	687.7	685.2	647.1	687.6	669.0	612.4
1946	644.8	497.3	723.9	780.3	688.7	675.1	742.4	862.9	825.9	945.8	873.5	813.6
1947	956.4	896.7	967.7	978.7	969.3	919.9	810.1	837.2	909.6	1021.6	913.6	954.8
1948	958.6	916.2	1052.2	948.9	884.4	956.9	817.3	925.8	963.9	1018.5	981.3	978.8
1949	921.4	876.7	965.7	851.8	790.4	822.1	614.8	779.9	779.9	617.1	627.0	784.1
1950	844.0	800.7	905.3	901.9	988.5	1023.5	864.8	1072.9	1043.2	1141.7	1049.2	1079.4
1951	1160.3	1054.8	1222.1	1160.5	1189.2	1122.3	895.2	1052.7	955.0	1112.6	1009.4	892.9
1952	1028.4	1002.8	1005.9	1026.8	929.6	680.4	483.1	831.0	946.9	1048.3	909.3	1003.9
1953	1020.8	994.6	1092.2	1091.6	1013.9	1021.9	901.3	905.2	938.3	957.6	848.8	892.8
1954	793.4	793.3	873.2	811.3								

Source: Bureau of the Census releases listed in Table 2. General adjustments made to the data as indicated in the accompanying text. Specific revisions as follows: Aluminum—Published figures for 1942 and 1943 were inflated to allow for the category "all other" not reported prior to 1944, the importance of this classification was assumed to be the same as it was to total shipments in 1944, 0.5 per cent. Copper—From July, 1944 to December, 1945, inclusive, toll operations and railroad captive foundry operations were excluded from the published information, reported figures increased by 15 per cent based on study of the importance of this segment of the industry by the Office of Price Administration [10]. Malleable iron—Prior to full coverage of the industry begun in 1943, two samples were employed, one from 1923 to 1930 and the other from 1930 to 1942. Degree of coverage of the later series was estimated by the Census on the basis of the Census of Manufactures; full coverage for the earlier series was approximated from a comparison of the figures reported from both samples in 1930. On this basis (the 1923–1930 sample representing 92.5 per cent of the total industry; the 1930–1942 sample; 88.0 per cent, 1930 to 1938 and 93.0 per cent, 1939 to 1942) the author inflated the published figures to full coverage. Steel—From January, 1926 to June, 1943 no shipment figures were published. The closest available series, Production of Commercial Steel Castings, was inflated to approximate full coverage by the author on the assumption that the Production of Commercial Steel Castings represented 80 per cent of the *reported* Shipments of All Steel Castings as they did in the period, July, 1943 to December, 1943, and that the *reported* figures, in turn, represented 95 per cent of full industry activity as was estimated by the Census. Figures for 1944 were inflated to full coverage by the author on the basis of the estimate by Census that the published figures represented 95 per cent of full coverage. Gray iron—Monthly figures for 1943 and 1944 were inflated to full coverage by the author on the estimate by the Census that reported coverage represented 93 per cent of full coverage.

3. SEASONAL ADJUSTMENTS

The data in Table 3 show a seasonal pattern. Relatively low rates of production during summer months, even during the war years, reflect the difficulty of obtaining high productivity in foundries during warm weather. These recurrent variations in output have been removed in two steps; each series has been adjusted for the number of working days in the month and for seasonal fluctuations.

3.1. Calendar Factors

The working schedule of foundries varied widely during this period; in general, it involved a single shift, a five-day week, and the observance of six holidays. No modification was made in the figures for variations in number of shifts; even during the war years substantially less than one-half of the foundries worked more than one shift [14, p. 17]. Further, adjustments for shift operations made in response to changing business conditions, even if they were possible, would tend to reduce the sensitiveness of the index to cyclical movements.

The first adjustment made in the figures is for changes in the number of days worked per month. In general, foundries operated five days per week throughout the period. No corrections were made for variations in working schedules which occurred as a result of changing business conditions in the early thirties or in the 1949 recession. Such corrections tend to make the index less sensitive to cyclical fluctuations, because alterations in the usual work week are, themselves, measures of business change. Moreover, adjustments of working schedules differed too radically from foundry industry to foundry industry and from area to area to render frequent modification valid. Exceptions to this are the period from 1923 to 1931, when the five and one-half day week was the custom and the period of World War II from January, 1941 to August, 1945, when foundries usually operated six days per week [13, pp. 54–55].

The second adjustment is for changes in the number of holidays. Prior to 1942, five holidays were observed: New Year's Day, Independence Day, Labor Day, Thanksgiving Day, and Christmas Day. During 1942, 1943, and 1944 only New Year's Day, Independence Day, and Christmas Day were celebrated; in 1945 Thanksgiving Day was added; and, in 1946 and later years, Memorial Day and Labor Day were included. On the basis of these work weeks and holidays, calendar adjustments in the monthly figures were made in order to remove the influence of variations in the days worked during the month.

3.2. Seasonal Factors

Pronounced seasonal changes arise from natural working conditions of the plants and fluctuating demands for their products. Since concurrent information is available for all metals from January, 1943, the ten-year period subsequent to this date was selected for a provisional computation of seasonal factors. The twelve-month-moving-average method of adjustment was employed, the resulting indexes were computed by use of the modified-mean technique whereby the extreme values found for each month were excluded from the calculation of the mean.[1] This method was useful in removing the influence of such incidental variations as the steel strike of July, 1952 from the final indexes.

A second series of seasonal indexes was computed, based this time on the six post-World-War II years of 1947 through 1952. These factors (Table 4) show a more pronounced pattern of seasonal trend than those based on a ten-year period including the war years of 1943, 1944, and 1945, and the post-war readjustment year, 1946. This is consistent with experience in other industries reported by the Federal Reserve Board [3, p. 1263].

The ferrous series, malleable iron, steel, and gray iron, have higher seasonal peaks in the spring than in the fall; in this they follow the pattern of durable manufactures in general. On the other hand, the nonferrous indexes, aluminum and copper, behave similarly to the producers of nondurable goods and reach their annual production heights in the fall.[2] This contrasting situation may be explained in part by the differing utilization of ferrous and nonferrous castings. Ferrous castings are used extensively in heavy construction and transportation machinery, industries showing high rates of activity in the spring; nonferrous castings are employed in aircraft, instruments, hardware, and light tools, industries having little seasonal fluctuations or high production levels in the fall [3, pp. 1292–3].

4. MOVEMENTS IN THE SERIES

After adjustment for calendar and seasonal fluctuations, the foundry series exhibit broadly similar patterns of output change as they move from phase to phase of the business cycle. Variations are still present, however, in the individual responses of the industries to different industrial situations.

---

[1] This method was utilized because fairly reliable information could be obtained from only a limited amount of material. Had information covering a longer period been available, the technique used by the Board of Governors of the Federal Reserve System would have been employed [1].

[2] In the newly revised Index of Industrial Production a single set of seasonal adjustment factors is applied to all primary metals [3, p. 1264].

4.1. *Relationships Among Foundry Series, 1943–1954*

The concurrent information available for all series shown in Figure 1, has been adjusted for seasonal fluctuations (Table 4) and stated comparably in pounds on an average-daily-shipments base. Most of the fluctuations can be identified with nationally-publicized events. Variations in the series during 1943 and 1944 resulted from two causes:

TABLE 4

SEASONAL ADJUSTMENT FACTORS OF SHIPMENT DATA OF SELECTED CASTINGS INDUSTRIES IN THE UNITED STATES, BASED ON 12-MONTH MOVING AVERAGE ADJUSTMENTS FOR 6-YEAR PERIOD, JANUARY, 1947, TO DECEMBER, 1952, INCLUSIVE

Month	Aluminum	Copper	Malleable	Steel	Gray Iron
January	99	101	101	98	103
February	105	104	104	107	103
March	103	105	103	106	105
April	104	100	105	103	105
May	101	102	104	102	101
June	97	95	101	104	100
July	80	82	84	82	85
August	91	95	92	90	93
September	108	106	104	102	103
October	108	102	101	100	101
November	106	106	102	102	102
December	98	102	99	104	99

Source: Table 3 and accompanying text.

(1) Adjustments were made to the data for all years on the basis of seasonal factors computed from the years 1947 to 1952, inclusive, a period in which seasonal variations were larger than during the war. The use of the same seasonal factors throughout stresses the mechanical character of the foundry index and the changes in seasonals which have occurred under specific conditions in the past. At the same time, these factors distort the adjusted series during 1943 and 1944 by introducing counter-seasonal variations. (2) Because controls were handled on a calendar-quarter basis by the War Production Board, they seem to have imparted some special quarterly fluctuations to the output of foundries. Apparently, more optimistic estimates at the beginning of a quarter resulted in larger allocations during the first month than in subsequent months, when more limited stocks appeared to be avail-

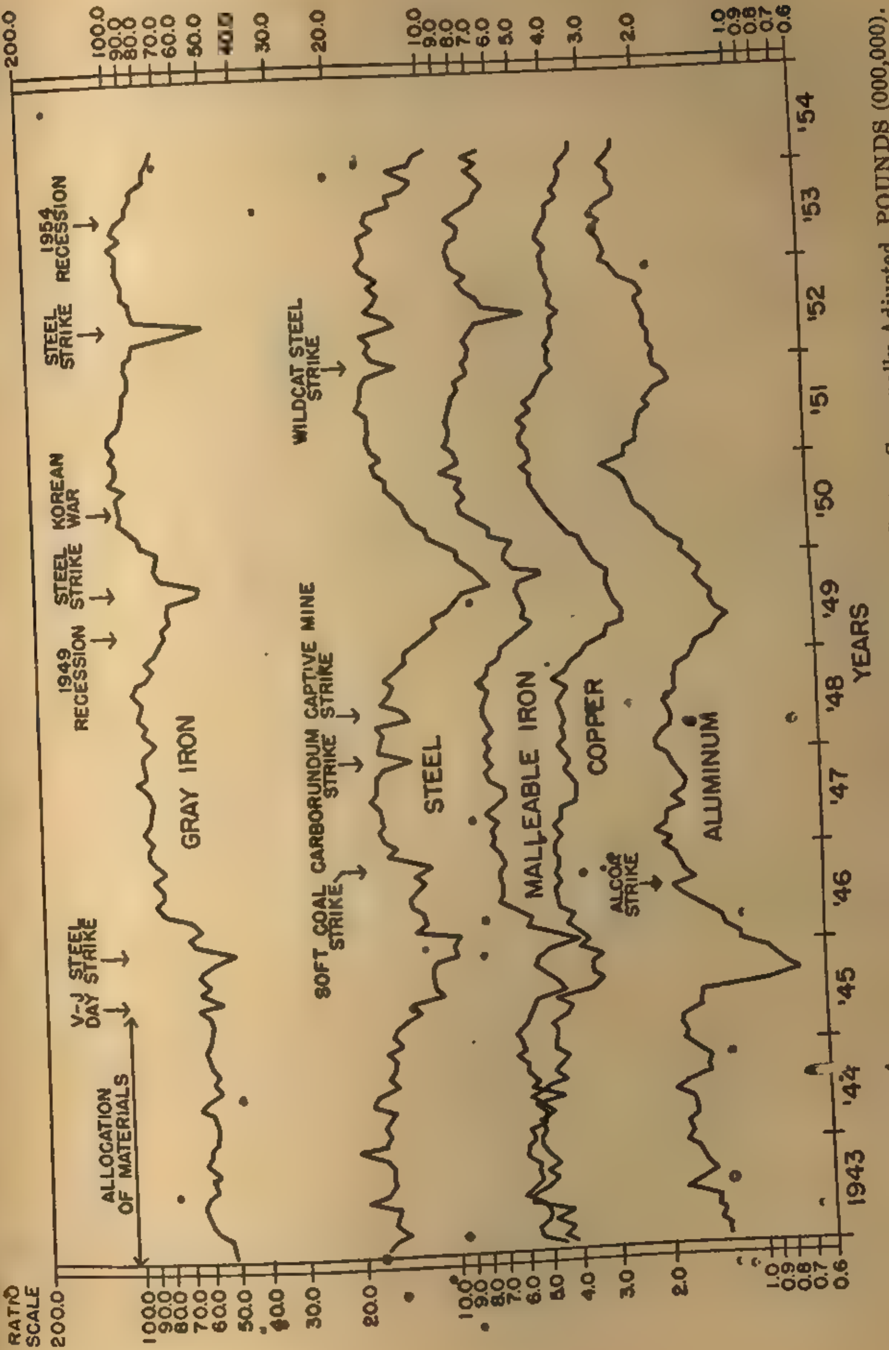


FIG. 1. Average Daily Shipments of Selected Castings in the United States, Seasonally Adjusted. POUNDS (000,000).

Source: Tables 3 and 4 converted to average daily shipments in pounds.

able; more pessimistic estimates tended to produce the reverse effect [13, p. 51].

In later years reconversion, strikes, and recessions were important influences. All of the series dropped in August, 1945, as the result of V-J Day; the more radical decline and slower recovery of the aluminum, copper, and steel series indicate a greater dependence on war work, especially in the case of aluminum. Except for steel, all industries declined in November, 1948 at the start of the 1949 recession. The lag of one month by steel may be accounted for by the dislocation of production schedules by strikes earlier in the year. General consistency of timing is notable also in the recovery from the 1949 recession; in June an upturn occurred in each series. The impact of the long steel strike during the fall makes it difficult to determine whether recovery would have been continuous or faltering had this interruption not taken place. Strikes in the steel industry usually have an immediate and pronounced effect on all of the ferrous foundries; strikes in associated industries such as coal and carborundum seem to have confined their effects to steel castings. Similarities in timing are evident also in the slight advances in each series following the outbreak of the Korean War and the downturns in each series at the beginning of the 1954 recession in August, 1953.

4.2. Relationships Among Foundry Series, 1929–1947

Since no monthly information concerning three of the series is available for the cyclically important pre-war period, less satisfactory annual data are employed to assess the magnitude of their reactions to severe depression and recovery. Production indexes based on figures from various *Census of Manufactures* [11] are shown in Table 5. In general, the industry coverages represented by these figures are similar to, but not identical with, the coverages involved in the monthly series.

Broadly speaking, the conclusions reached earlier concerning the behavior of the casting series are reinforced by this information. In general, the production changes of all foundry industries show wide cyclical movements; again, individual dissimilarities are apparent. In 1937, recoveries from depression lows were much more complete in the ferrous than in the nonferrous series, but in 1939, however, output declined in all fields except aluminum where the beginnings of expanded aircraft production were reflected.

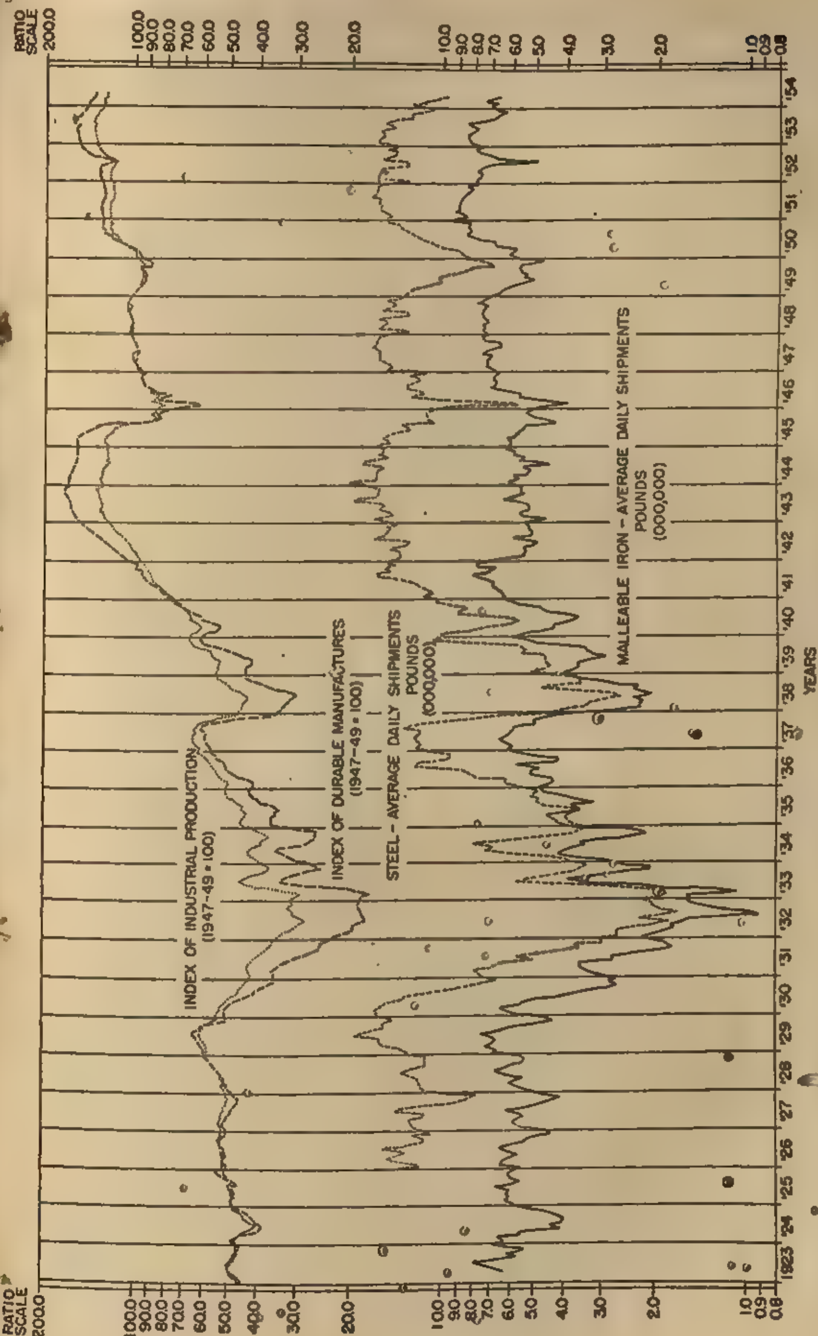


FIG. 2. Comparison of Indexes of Industrial Production and Durable Manufactures with Average Daily Shipments of Malleable-Iron and Steel Castings in the United States, 1923 to 1954, Inclusive, SEASONALLY ADJUSTED

Source: Malleable iron and steel series from Table 3, seasonally adjusted with factors from Table 4, and converted to average daily shipments in pounds. Federal Reserve Indexes from [3, pp. 1324 and 1326].

5. COMPARISON WITH INDEXES OF INDUSTRIAL PRODUCTION

In Figure 2 the longer record of operations in the malleable iron and steel fields is compared with the general Index of Industrial Production and the component Index of Durable Manufactures. This permits an examination in detail of changes in output of at least two of the foundry industries during a period of wide cyclical movements, information not obtainable directly from the limited historical series available for aluminum, copper, and gray-iron.

TABLE 5

INDEXES OF PHYSICAL PRODUCTION: ALUMINUM, COPPER, MALLEABLE-IRON, STEEL, AND GRAY-IRON CASTINGS IN THE UNITED STATES FOR SELECTED YEARS, 1929 TO 1939, INCLUSIVE

1929 = 100

Year	Metal				
	Aluminum	Copper	Malleable Iron	Steel	Gray Iron
1929	100	100	100	100	100
1931	48	51	41	33	47
1933	33	18	36	21	37
1935	58	24	59	37	55
1937	56	58	90	86	94
1939	79	43	69	50	79

Source: [11: *1931*, pp. 839, 880, 908, and 992; *1935*, p. 1079; *1937*, Pt. I, pp. 936, 1023, and 1061; *1939*, Vol. II, Pt. 2, pp. 198, 202, 26?, 342, and 347].

5.1. *Foundry Data More Volatile*

Three differences between the foundry series and the indexes of the Federal Reserve Board account for most of the contrasting reactions: (1) When compared with the larger and broader indexes the foundry information is, in a sense, a small sample of business activity, as such it tends to display the well known larger sampling errors of all small samples. (2) The casting series are computed more mechanically than the Federal Reserve Board's indexes. With foundries a simple procedure has been followed by applying the same corrections and adjustments throughout the period; for example, seasonal factors computed from the years 1947 to 1952 have been used from the beginning to the end of the period. In the Federal Reserve Board's indexes seasonal

adjustments are modified continually and estimates are made before all returns are received; these practices tend to introduce subjective evaluations into the preliminary estimates which may be carried over to some extent in the final or revised figures. (3). In the foundry series the utilization throughout the period of a single set of seasonal adjustment factors, based on the 1947–1952 period, underadjusts some important seasonal movements prior to World War II. This is especially notable in the malleable-iron data, for the seasonal requirements of the manufacturers of automobiles, railroad equipment, and heavy construction machinery were much greater in the 1920's than they are now [3, pp. 1292–3 and 2, pp. 2–4].

5.2. Foundry Data Show Little Secular Trend

The general level of foundry activity has changed only a small amount in the past thirty years. The shipments of steel castings in 1929 were exceeded by less than 8 per cent in the boom years of World War II; the output of malleable iron expanded only about one-third between 1929 and the Korean War. On the other hand, the Index of Industrial Production showed an increase of more than 100 per cent, the Index of Durable Manufacturers, 150 per cent. This characteristic of foundry data may be explained by the persistent shift from castings to other methods of forming metal parts, such as stamping, forging, and welding.

5.3. Foundry Information More Sensitive to Cyclical Change

The large utilization of castings for capital improvements is reflected in the amplitude of changes in output levels during business fluctuations. In fact, the relative volatility of each of these series seems to vary with the importance of capital goods in the index. Thus, the foundry series are more sensitive than the Durable Index, while the latter, in turn, is more variable than Industrial Production Index.

As the period advanced, the proportionate contractions from peak to trough (1929–1932, 1937–1938, and 1948–1949) became progressively greater in the foundry series than in the Federal Reserve indexes; thus, the declines for malleable iron and steel from 1948 to 1949 were 40 per cent and 58 per cent, respectively, while they were only 10 per cent and 17 per cent for Industrial Production and Durable Manufactures. This development may be explained by the new system of weights being employed by the Federal Reserve Board; instead of basing the importance of an industry on the gross value of its product, its influence is determined by the value *added* to its product. This. of

course, decreases the significance of industries in which the value added to the product is small in proportion to the total value of the product, and increases the emphasis of industries with a high value-added ratio. The importance of the change in weights here is that it now renders the Federal Reserve indexes less sensitive to cyclical change than they were before the revision, since items with high value-added ratios such as tools, instruments, and machinery are generally less sensitive to business change than goods with low value-added ratios such as coal, lumber, and blast furnace products.[3] [Cf. 3, p. 1243 and 1277].

5.4. Foundry Information Changes in Phase With Other Indexes

In the foundry series it is difficult to distinguish the actual peaks and troughs of business activity from the residual seasonal fluctuations and the incidental changes resulting from the small character of the sample. Apart from these problems, once cyclical changes are located in the Index of Industrial Production, similar movements are apparent in the malleable and steel series. Since the Index exhibits a slight lead at peaks and a rough coincidence at troughs [9, p. 60], it may be assumed that the foundry series do not depart markedly from this timing.

6. AN INDEX OF FOUNDRY ACTIVITY

Although foundry series generally move at the same time and in the same direction, they do exhibit individual variations that tend to distract attention from the more fundamental responses all foundries make to economic change. An effective way of overcoming this difficulty is to combine the five casting series into a composite index of foundry activity.

6.1. Alternative Weighting of Foundry Index

Production relatives with a base 1947–49 were computed for each foundry series shown in Table 3 after the data had been adjusted for seasonal variations (Table 4). The resulting schedules were combined into four composite indexes of foundry activity; the first was weighted equally, the second by physical volume of shipments, the third by dollar volume of shipments, and the fourth by value added to product. The weights were derived from the data in Table 1.

Each of these devices has certain unique advantages: (1) By the system of equal weights the composite index reflects changes in each

[3] Ex. the low point reached in the fall of 1949 by the Metal Fabricating component of the Index of Industrial Production was only 80, while the comparable point for the Primary Metals component was 46 [3, p. 1298].

of the five industries without regard to the physical or monetary importance of the castings shipped. (2) The method of weighting the series according to the number of pounds of castings shipped by each industry places greatest emphasis on the gray-iron group. These foundries were found to exhibit greater production stability than the other series because their castings are used more extensively in consumers' products than they are in war or defense lines.

(3) The technique of weighting the series according to the value of the castings produced by each industry results in giving less importance to the gray-iron group and somewhat more to the costly nonferrous castings. It imparts emphasis to aircraft and marine castings, prominent in defense and war activity. (4) The use of value added data as a basis of weights is consistent with the weighting system used in the Indexes of Industrial Production and Durable Manufactures. In effect, this weighting reduces the relative importance of nonferrous castings by eliminating the costly metal from the figures. Actually, the index based on value added to product is almost identical with the one based on total value of product.

6.2. Selection of the Most Appropriate Weights

It is not enough to come to a logical conclusion in the selection of weights; it is necessary, also, to see how well the index employing the weights behaves in practice. Two sets of data are available: (1) biennial figures for the period, 1929–1939, and (2) annual information for the period, 1943–1953.

6.2.1. The pre-war period. Shown in Table 6 are weighted indexes of foundry activity derived from the Census of Manufactures data for the years 1929 through 1939. The relatives, on a 1929 base (Table 5), have been combined and weighted by the importance of each metal in 1947 (Table 1) to form indexes of foundry activity similar to those described above. For comparison, five related indexes of general economic activity are presented: (1) Private Investment, a narrow measure which includes: construction of new plants by business, purchases of producers' durable equipment, and changes in business inventories. It is, in effect, a measure of domestic, non-government investment. (2) Total Investment, a broader series which adds to (1) above: net foreign investment (current balance of payments) and government purchases of goods and services. This is consistent with the usual definition of investment. (3) Durable Manufactures and Industrial Production, current Federal Reserve indexes adjusted to the 1929 base. (4) Gross National Product, a measure of the total economic activity

in the United States. This series and the two investment ones were computed from data stated in constant (1939) dollars [15].

A comparison of the related indexes in Table 6 reveals a general similarity of behavior, with two important exceptions: (1) Private Investment, Total Investment, and Durable Manufactures reacted more violently to the depression than the other two series; this observation is consistent with the notion that investment and durable goods activity change more drastically than general business operations during

TABLE 6

INDEXES OF FOUNDRY ACTIVITY WEIGHTED VARIOUSLY WITH 1947 FACTORS COMPARED WITH INDEXES OF RELATED ACTIVITIES IN THE UNITED STATES FOR SELECTED YEARS 1929 TO 1939, INCLUSIVE

(1929 = 100)

Year	Foundry Indexes Weighted				Related Indexes				
	Equally	Physical Volume	Dollar Volume	Value added	Private Investment	Total Investment	Durable Manufactures	Industrial Production	Gross National Product
1929	100	100	100	100	100	100	100	100	100
1931	44	46	45	45	40	46	52	68	84
1933	29	35	32	33	11	31	40	63	72
1935	47	53	49	50	45	50	63	80	86
1937	77	91	85	86	77	68	92	103	102
1939	64	[illegible]	70	70	66	71	82	98	106

Source: Tables 1 and 5; [15, pp. 26–27]; and [3, pp. 1324 and 1326].

depressions. (2) From 1937 to 1939 Total Investment and Gross National Product increased while the other series decreased; evidently spending by the Federal government, which increased Total Investment and was reflected in a larger Gross National Product, did not carry over to Private Investment, Durable Manufactures, or Industrial Production sufficiently to expand the series. Fluctuations in each of the foundry indexes follow movements in Private Investment.

6.2.2. The war and post-war period. In Table 7 indexes similar to those shown in Table 6 are presented for a later period and computed to a later base, 1947–1949. The foundry indexes are derived from the monthly values in Table 3; the related indexes are from the national income[4] and Federal Reserve data used in computing the comparable series in Table 6.

[4] In this period, quarterly national income information is available on a current dollar basis; it was not employed because the precision added by quarterly figures did not appear to be as great as inherent inflationary variations which could not be completely removed from the data.

Substantially more diversity is evident in the movements of the related series during this, than during the earlier period. The enormous expenditures by the government for war equipment, especially in 1943, 1944, and 1945, account for a large portion of the differences. Conflicting changes which occurred during the later years are, however, a logical outcome of the economic process. Castings are employed in the construction of plants and equipment for the manufacture of pro-

TABLE 7

INDEXES OF FOUNDRY ACTIVITY WEIGHTED VARIOUSLY WITH 1947 FACTORS COMPARED WITH INDEXES OF RELATED ACTIVITIES IN THE UNITED STATES, 1943–1953, INCLUSIVE (1947–49 = 100)

Year	Foundry Indexes Weighted				Related Indexes				
	Equally	Physical Volume	Dollar Volume	Value Added	Private Investment	Total Investment	Durable Manufactures	Industrial Production	Gross National Product
1943	101	77	87	86	27	163	162	127	103
1944	105	77	89	88	33	183	159	125	110
1945	89	75	81	80	42	162	123	107	108
1946	93	83	88	87	102	103	86	90	97
1947	106	100	102	102	96	97	101	100	98
1948	109	102	104	104	114	105	104	104	101
1949	82	82	82	82	90	98	95	97	101
1950	110	104	107	106	134	114	116	112	110
1951	124	117	120	119	138	141	128	120	118
1952	110	100	103	103	122	146	136	124	121
1953	119	106	111	110			153	134	

Source: Tables 1 and 3; [15, pp. 26–27]; [3, pp. 1324 and 1326]; and *Federal Reserve Bulletin*, March, 1954, p. 295, for "Preliminary" 1953 values for Index of Industrial Production and Index of Durable Manufactures.

ducers' goods, consumers' goods, and war goods. The outbreak of war in Korea caused an initial expansion of Private Investment and concomitant foundry production to create and adapt new factories before large government orders could be filled. Thus, increases in government spending to swell Total Investment were preceded by the utilization of private investment funds by foundries and others in building or modernizing facilities with which to make the proper war goods.[5] Contrasting movements of Private and Total Investment from 1951 to 1952 emphasize the curtailing of Private Investment (and foundry activity) after the production facilities for the new war were developed.

[5] Actually the Federal government's total purchases of goods and services were larger by $2.1 billion (1939 dollars) in 1949 than they were in 1950 [15, p. 27].

On the basis of these comparisons, weighting the foundry series by value-added figures seems to be most appropriate: (1) It gives each foundry industry and associated area an importance relative to the expenses of producing the castings without regard to the cost of the metal from which the castings are formed. (2) It tends to give less influence to nonferrous castings than the equally-weighted scheme and thus provides an index less heavily affected by occurrences in the aircraft and marine fields. (3) It gives less prominence to gray-iron castings and more to aluminum than the physically-weighted index and so tends to present a more balanced picture of the over-all investment economy. (4) It is, in essence, a workable average system of weighting which possesses the good points of all of the foundry series without the vulnerability of either of the extreme weights. Values for the Index of Foundry Activity are shown in Table 8.

7. EVALUATION OF THE INDEX

The appraisal of the Index of Foundry Activity is facilitated by a comparison with the Indexes of Industrial Production, Durable Manufactures, and Private Investment in Figure 3. Specific attributes are apparent.

7.1. Timing of Movements Roughly Similar to Those in Index of Industrial Production

Similarities of movement are evident between the Foundry Index and the Index of Industrial Production; since the latter has been found to exhibit some lead at peaks and general consistency at troughs [9, p. 60], it may be reasoned that the Foundry Index, also, is timed in this general fashion.

7.2. Magnitude of Movements Greater than Index of Industrial Production

Over the period investigated, the size of the fluctuations in foundry activity has been substantially greater than changes in the Index of Industrial Production and Gross National Product. In other than war years, changes in the Foundry Index approximate in a rough way the amplitude of movements of private investment (Tables 6 and 7); in spite of the difficulty of comparing annual and monthly data in Figure 3, the generalization is supported also by this illustration.

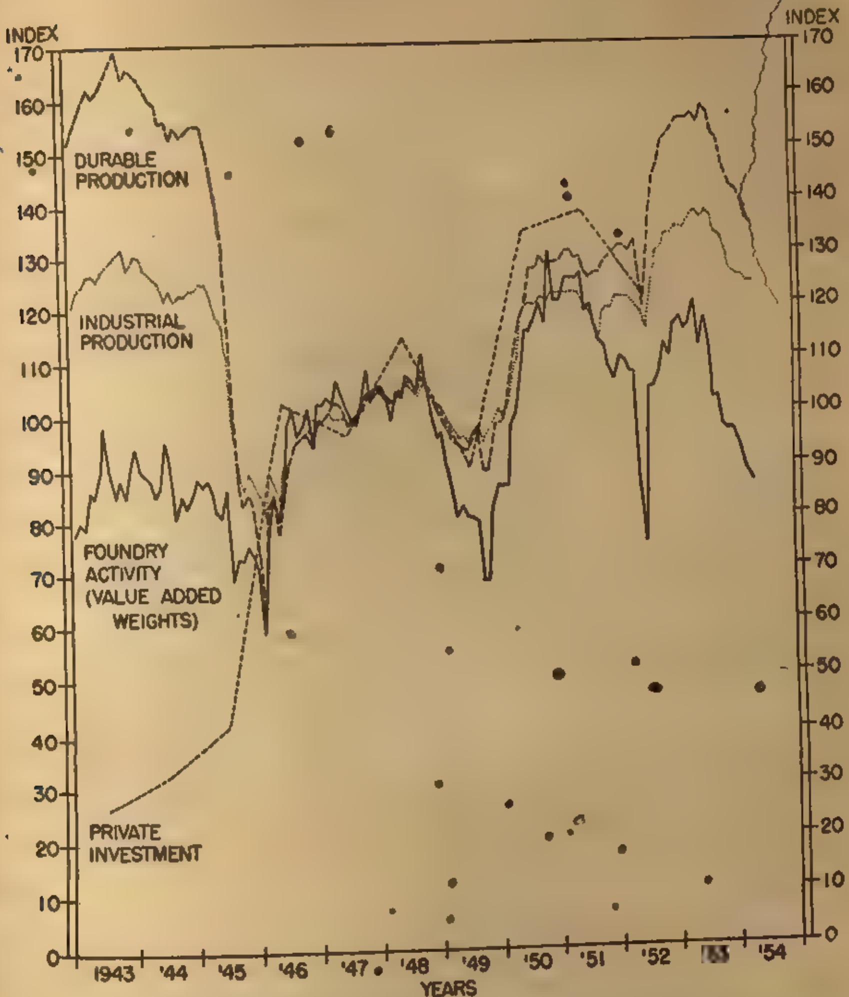


FIG. 3. Comparison of Index of Foundry Activity with Indexes of Industrial Production, Durable Manufactures, and Private Investment in the United States, 1943 to 1954, Inclusive, SEASONALLY ADJUSTED. (1947–49 = 100).

Source: Foundry data from Table 3, seasonally adjusted by factors from Table 4, and combined in a weighted-relative index using the value-added-to-product figures from Table 1, Table 8, and [3, pp. 1324 and 1326].

TABLE 8

INDEX OF FOUNDRY ACTIVITY IN THE UNITED STATES SEASONALLY ADJUSTED

(1947–49 = 100)

Year	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1943	78	80	79	86	85	90	98	90	85	88	85	90
1944	94	90	89	88	85	87	95	91	81	85	83	86
1945	88	87	88	86	82	81	86	69	73	73	75	74
1946	71	59	78	85	80	86	98	101	96	98	101	94
1947	102	102	103	102	106	103	101	99	98	102	108	102
1948	105	105	103	99	104	103	107	106	104	111	105	100
1949	95	96	89	85	80	82	80	80	79	68	68	82
1950	86	86	86	97	100	106	115	115	117	120	117	130
1951	121	121	125	125	125	126	119	120	116	113	112	108
1952	106	110	109	107	107	87	75	104	104	107	112	110
1953	116	117	115	116	120	112	117	112	102	102	97	96
1954	96	93	88	86								

Source: Foundry data from Table 3, seasonally adjusted by factors from Table 4, and combined in a weighted-relative index using the value-added-to-product figures from Table 1.

7.3. *Determination of Index Values Highly Mechanical*

Index values may be obtained by applying a stereotyped set of procedures to the raw data. A disadvantage of this method is the transmitting of local, extraneous variations to the final index. Two alternatives are available; (1) identifying the special events by supplementary notes, and (2) editing the original information to remove the extraneous material. The first alternative has been selected here on the assumption that some incidental fluctuations and possible misinterpretations of the index are easier to comprehend and to correct than an unknown amount of editing.[a]

7.4. *Availability of Data Relatively Prompt*

Foundry data in the Facts for Industry series are customarily issued about six or seven weeks after the end of the month for which information is published. This is about as soon as the Preliminary *Index of Industrial Production* is released by the Board of Governors in the mimeographed Business Indexes supplement to *National Summary of Business Conditions*. It is substantially earlier than private investment data collected by the Office of Business Economics of the Department of Commerce and the Securities and Exchange Commission are published in the *Survey of Current Business* [5, p. 9].

REFERENCES

[1] Barton, H. C., Jr., "Adjustment for seasonal variation," *Federal Reserve Bulletin*, 27 (1941), 518–28.
[2] Board of Governors of the Federal Reserve System, *Seasonal Adjustment Factors*, Washington (1940).
[3] Board of Governors of the Federal Reserve System, "Federal Reserve monthly index of industrial production," *Federal Reserve Bulletin*, 39 (1953), 1239–1328.
[4] Clark, J. M., *The Economics of Overhead Cost*, Chicago (1923).
[5] Foss, M. A., "Investment programs and sales expectations in 1952," *Survey of Current Business*, 34 (1954), 9–12.
[6] Haberler, G. von, *Prosperity and Depression*, Geneva (1939).
[7] Hussey, Miriam, *Foundry Activity as a Business Barometer*, Philadelphia (1950).
[8] Mitchell, W. C., *What Happens During Business Cycles*, New York (1951).
[9] Moore, G. H. *Statistical Indicators of Cyclical Revivals and Recessions*, New York (1950).
[10] Office of Price Administration, *Statement of Considerations Involved in the Issuance of Amendment No. 8 to Revised Maximum Price Regulation No. 125 —Nonferrous Foundry Products*, Washington (1945)
[11] U. S. Bureau of the Census, *Biennial Census of Manufacturers. 1931, 1935, 1937*, and *1939*, Washington (1935, 1938, 1939, and 1942).
[12] U. S. Bureau of the Census, *Census of Manufactures*: 1947. Washington (1949).
[13] U. S. Bureau of Labor Statistics, *Handbook of Labor Statistics*, Washington (1948). (1947 Edition).
[14] U. S. Bureau of Labor Statistics, *Wage Structure: Foundries, 1945*, Washington (1946).
[15] U. S. Department of Commerce, "National income and product of the United States, 1952," *Survey of Current Business*, 33 (1953), 6–32.
[16] War Production Board, *Wartime Production Achievements*, Washington, (1945).

CARGO LOSS IN FERRYING OPERATIONS

WALTER L. DEEMER, JR.
*United States Air Force**

1. INTRODUCTION

In ferrying operations of valuable items (e.g., aircraft spare parts needed in a theater of operations during a war), the number of items to be carried in each aircraft is frequently under the control of the planner. By using more aircraft he can make the loading per aircraft smaller and hence possibly reduce the probability of large losses at the risk of increasing the probability of small losses. (The small loading also increases the actual ferrying cost.) Because of the value of the items it may be a good investment to buy this insurance.

In making his decision as to how many items to load per aircraft, the planner therefore needs to know, as a function of the loading, the probability of losing any given number of items. Sometimes the mean and variance may give sufficient information, without knowledge of the more laboriously calculated probabilities. Usually however, the distributions are not even approximately symmterical nor even unimodal, so that an adequate approximation using only the mean and variance is not possible.

In this paper expressions are given for the mean and the variance of the number of items lost, for the probability of losing a given number of items and for the probability generating functions. The moment generating functions may be found from the probability generating functions by substituting e^ϕ for ϕ. For models 1, 3, and 4 several formulas, in addition to the probability generating function and the general formula (m.4) are given for the probabilities of losing exactly a items.[1] These special formulas are useful for certain values of a. Apparently this problem has not been previously considered in this form in the statistical or operational research literature.

We denote by k the number of flights to be used in the ferrying operation. No assumptions are required as to how many aircraft are used in the operations; there may be k aircraft, each making one trip, or fewer aircraft may be used, some or all making more than one trip.

Each aircraft carries r valuable items, so that a total of $N=kr$ items are ferried. Each item is assumed to be equipped with a life raft or some other means of preventing it from sinking in case the aircraft

* The views expressed here are those of the author and are not to be construed as reporting official or unofficial policies of the United States Air Force.

[1] ($m.x$), ($m=1, 2, 3, 4$) represents formula x for Model m. These are exhibited in Table 1.

ditches. In the event that the aircraft ditches it is assumed that the items leave the aircraft and float. (If the objects are inanimate, it is assumed that they are jettisoned.) Some or all of the life rafts or flotation devices may be equipped with radio transmitters or radar reflectors to help the search.

It is assumed that there is a constant probability p that an aircraft will be forced to ditch (i.e., make a forced landing on the water) before it has delivered its items. It is also assumed that each trip is independent of all the other trips, so that the distribution of number of aircraft ditched is given by the expansion of $(p+q)^k$, where $q=1-p$.

When an aircraft is ditched, a search is made for the floating objects.[2] Some of these may be recovered. The items not recovered will be referred to as lost.

Four models are considered, each model being defined by a set of assumptions on the behavior of the floating items and the characteristics of the search operation.

2. NOTATION

k = number of carriers used

r = number of items per carrier

$N=kr$ = total number of items carried

p = probability that a carrier will ditch

$q=1-p$

t = probability that an individual item will be found

$s=1-t$

f = probability that a clump of lost items will be located

$g=1-f$

$$A(k, p, j)=C_j^k p^j q^{k-j}=[k!/j!(k-j)!]p^j q^{k-j}, \text{ if } 0\leqq j\leqq k.$$
$$\equiv 0, \text{ if } j>k \text{ or } j<0.$$
$$A(0, p, 0)\equiv 1$$
$$D_t^a=\partial^a/\partial t^a$$

$E_m(X)$ = expected value of the random variable X under model m

$V_m(X)$ = variance of X under model m.

3. DESCRIPTION OF THE MODELS

Model 1. By the assumptions of this model the ditched items (jr in number if j aircraft are ditched) are distributed over a wide area so that the conditional distribution of number of items recovered, given that j aircraft have ditched, is given by the terms of $(t+s)^{jr}$ where t is the probability of recovering a single item; $s=1-t$.

[2] The cost of maintaining search facilities is not relevant to this problem because these are maintained in any case for the rescue of crews.

Model 2. Here the items are assumed to float together (either because they are tied together or because the wind and waves have not separated them) so that if one item is found they are all found. Then if f is the probability of finding a clump of r items, the conditional probability of finding ir items given j aircraft are ditched is the term in f^i of the expansion of $(f+g)^j$, where $g=1-f$ and as before j is the number of aircraft ditched.

Model 3. Here the items do not clump to the extent they do in Model 2. We assume that there is a probability f of getting in the vicinity of the r semi-dispersed items and then having arrived in the vicinity of the items there is a probability t of finding a single item. (One might get in the vicinity of a clump, for example, by finding the ditched aircraft or debris from it.) The conditional probability of getting in the vicinity of i clumps given that j aircraft have been ditched is the term in f^i of $(f+g)^j$. For each such clump the probability of recovering a items is the term in t^a of $(t+s)^r$.

Model 4. This is like Model 3 except it is assumed that the way one gets in the vicinity of a clump is actually to find an item. For example, one of the r items on each aircraft might have a special radio transmitter or a radar reflector. Here then the probability of finding the first item of a clump of r items is f. Having found the first item, the probability of finding each successive item in that clump is t.

General Comments on the Models. It is not claimed that the mathematical models here considered fit exactly any actual ferrying situation. For example, it is somewhat unlikely that any actual ferrying situation fits Model 1, though it might fit when the items are dropped individually by parachute over a wide area. Some of the other models seem to be adequate representations of actual situations.

Models 1 and 2 are special cases of Models 3 and 4 as follows: If $f=1$ in Model 3 we have Model 1. If $t=1$ in Model 3 or Model 4 we have Model 2.

4. RESULTS

The results are given in Table 1. In the summation indices $[(a-1)/r]$ means the greatest integer less than or equal to $(a-1)/r$; for $a=0$, this is equal to -1.

5. DISCUSSION OF RESULTS

In Models 1, 2, and 3 the expected number of items lost is not a function of the loading, as evidenced by the fact that k and r appear only as the product $kr=N$ in the expressions for the expected number

TABLE 1

SUMMARY OF RESULTS

	Model 1	
Expected number of items lost	$kprs$	(1.1)
Variance of number of items lost	$kprs(t+rsq)$	(1.2)
The probability generating function	$M_1(\phi)=[p(t+s\phi)^r+q]^k$	(1.3)
Formulas for probability of losing* exactly $a=br+c$ items $(0\leq c<r)$	$\sum_j A(k,p,j)A(jr,s,a)$	(1.4)
	$(s^a/a!)D_t^a(q+pt^r)^k$ (useful for small a)	(1.5)
	Model 2	
Expected number of items lost	$kprg$	(2.1)
Variance of number of items lost	$kpr^2g(1-gp)$	(2.2)
The probability generating function	$M_2(\phi)=[p(f+g\phi^r)+q]^k$	(2.3)
Formula for probability of losing exactly $a=br$ items	$p^bg^bC_b^k(q+pf)^{k-b}$	(2.4)
	Model 3	
Expected number of items lost	$kpr(1-tf)$	(3.1)
Variance of number of items lost	$kpr[qr(1-tf)+tf(s+rtg)]$	(3.2)
The probability generating function	$M_3(\phi)=\{p[f(t+s\phi)^r+g\phi^r]+q\}^k$	(3.3)
Formulas for probability of losing exactly $a=br+c$ items $(0\leq c<r)$*	$\sum_j \sum_{i=j-b}^{j} A(k,p,j)\,A(j,f,i)\,A(ir,t,jr-a)$	(3.4)
	$\sum_{n=0}^{b}\{C_n^kg^np^ns^{a-nr}/(a-nr)!\}\,D_t^{a-nr}(q+pft^r)^{k-n}$ (useful for small a)	(3.5)
	$\sum_j \{A(k,p,j)t^{jr-a}/(jr-a)!\}\,D_s^{jr-a}(g+fs^r)^j$ (useful for $a-kr$ small)	(3.6)
	Model 4	
Expected number of items lost	$kp[r-(s+rt)f]$	(4.1)
Variance of number of items lost	$kp\{fts(r-1)+fg(s+rt)^2+q[r-f(s+rt)]^2\}$	(4.2)
The probability generating function	$M_4(\phi)=\{p[f(t+s\phi)^{r-1}+g\phi^r]+q\}^k$	(4.3)
Formulas for probability of losing exactly $a=br+c$ items† $(0\leq c<r)$	$\sum_j\sum_i A(k,p,j)\,A(j,f,i)\,A[i(r-1),t,jr-i-a]$	(4.4)
	$\sum_{n=0}^{b} C_n^kg^np^ns^{a-nr}/[(a-nr)!]\,D_t^{a-nr}(q+pft^{r-1})^{k-n}$ (useful for small a)	(4.5)

* $\sum_j$ is the sum from $j=[(a-1)/r]+1$ to $j=k$.
† $\sum_i$ is the sum from $i=j-b$ to $i=\min(jr-a,j)$.

of items lost. In Model 4, the expected number of items lost is an increasing function of r, the loading per aircraft. This results from the fact that in getting in the vicinity of a clump one item is found. In all models the variance is an increasing function of r, the loading per aircraft.

After deciding (in conference with the operations officer) which model is most applicable to the actual situation, the analyst can prepare numerical tables showing the probability of losing a items for various loadings, based on estimates of the probabilities p, t, and f involved; p, t, and f can frequently be estimated rather reliably from previous experience. These tables can be used in deciding what loading to use.

The probability of losing a items has been found useful in making decisions as to the desirability of developing salvage equipment, and as to the best type of salvage equipment.

6. DISCUSSION OF FORMULAS

6.1. *The Probability Generating Functions*[3]

The functions $M_m(\phi)$ are the expected values of ϕ^X, where X is the random variable, number of items lost:

$$M(\phi) = P\{X = 0\} + \phi P\{X = 1\} + \cdots + \phi^{kr} P\{X = kr\}.$$

and hence the probability that $X=a$ is the coefficient of ϕ^a in the expansion of $M(\phi)$.

When either k or r is small, $M(\phi)$ is easy to expand and the numerical evaluation after the expansion is quite easy. When both k and r are large, the expansion and numerical evaluation are usually tedious.

Example

When $r=1$, the coefficient of ϕ^a in $M_3(\phi)$ is

$$C_a^k(pft + q)^{k-a}(pfs + gp)^a.$$

When $r=2$, $k=3$ the expansion of $M_3(\phi)$ is

$$\begin{aligned} M_3(\phi) = {} & A^3 + 3A^2B\phi + (3AB^2 + 3A^2C)\phi^2 \\ & + (B^3 + 6ABC)\phi^3 + (3B^2C + 3AC^2)\phi^4 \\ & + 3BC^2\phi^5 + C^3\phi^6, \end{aligned}$$

where: $A=pft^2+q$; $B=2\,pfts$; $C=pfs^2+gp$.

[3] For a more complete discussion of probability generating functions see W. Feller, *An Introduction to Probability Theory and Its Applications*, New York, John Wiley and Sons, 1950, p. 212 ff.

The advantage of $M_m(\phi)$ for calculating $P\{X=a\}$ depends not on the value of a but on the values of r and k. The other special formulas, discussed below, are suitable for special values of a, independent of r and k.

6.2. *Special Formulas for Large and Small a*

Whether the special formulas (1.5), (3.5), (3.6), and (4.5) are simpler than the general formulas (m.4) depends not on the value of r and k but on the value of a.

The utility of these special formulas may be illustrated by exhibiting the formulas for $P\{X=0\}$, $P\{X=1\}$, and $P\{X=6\}$ for $(k=6, r=1)$ using the equations (3.4), (3.5), and (3.6).

By (3.4)

$$\begin{aligned}
P\{X = 0\} &= A(0, f, 0)A(0, t, 0)A(6, p, 0) \\
&\quad + A(1, f, 1)A(1, t, 1)A(6, p, 1) + \cdots \\
&\quad + A(6, f, 6)A(6, t, 6)A(6, p, 6); \\
P\{X = 1\} &= A(6, p, 1)[A(1, f, 0)A(0, t, 0) + A(1, f, 1)A(1, t, 0)] \\
&\quad + A(6, p, 2)[A(2, f, 1)A(1, t, 1) + A(2, f, 2)A(2, t, 1)] \\
&\quad + \cdots \\
&\quad + A(6, p, 6)[A(6, f, 5)A(5, t, 5) + A(6, f, 6)A(6, t, 5)]; \\
P\{X = 6\} &= A(6, p, 6)[A(6, f, 0)A(0, t, 0) + A(6, f, 1)A(1, t, 0) \\
&\quad + \cdots + A(6, f, 6)A(6, t, 0)].
\end{aligned}$$

By (3.5)

$$P\{X = 0\} = (q + pft)^6;$$
$$P\{X = 1\} = 6p(q + pft)^5(sf + g).$$

By (3.6)

$$P\{X = 6\} = p^6(g + fs)^6.$$

These examples show the great simplification possible with the specialized formulas. On the other hand, the use of equations (m.4) may be systematized so that the routine of setting up a computing table and computing from it may be easily learned and for the average computer this is an advantage. In Section 8 an example is worked in detail and in Table 2 a sample computing sheet is shown.

7. OUTLINE OF PROOFS OF EQUATIONS

Three random variables will occur; these will be denoted as follows:

X: Number of items lost
Y: Number of aircraft ditched
Z: Number of clumps located.

The index m refers to the models. For example, $P_m\{X=a\}$ means the probability of losing a items under model m.

A perfectly general equation for $P_m\{X=a\}$ is

$$(7.1)\qquad \begin{aligned} P_m\{X = a\} = \sum_j \sum_i P_m\{X = a \mid Y = j; Z = i\} \\ \cdot P_m\{Z = i \mid Y = j\} \cdot P_m\{Y = j\}. \end{aligned}$$

Conceptually j runs from 0 to k and i runs from 0 to j. In the equations given in Table 1 for $P_m\{X=a\}$, limits on j and i which do not cover these ranges are given because for some values of i and j some of the probabilities are identically 0 and hence these terms are omitted. For example, in $P_m\{X=kr\}$ only $j=k$ gives a non-zero value of $P_m\{X=kr \mid Y=j; Z=i\}$. Equations ($m$.4) of Table 1 ($m=1, 2, 3, 4$) are based on (7.1) with the substitution for the conditional probabilities of the binomial terms appropriate to the particular model.

The probability generating functions (m.3) are the expected values of ϕ^X:

$$(7.2)\qquad E_m(\phi^X) = \sum_{a=0}^{kr} \phi^a P_m\{X = a\}.$$

These may be readily evaluated by using the right side of (7.1) for $P\{X=a\}$, first evaluating

$$\sum_{a=0}^{kr} \phi^a P_m\{X = a \mid Y = j; Z = i\},$$

then multiplying by $P_m\{Z=i \mid Y=j\}$ and summing over i and finally multiplying by $P_m\{Y=j\}$ and summing over j.

The probability generating functions may also be evaluated by formal manipulation of conditional expectations.[4] If X is a random variable with a binomial distribution: $P\{X=j\}=A(k, p, j)$, then

$$E(\phi^X) = (p\phi + q)^k.$$

The conditional expectation of X given Y and Z may be evaluated using this fact; the conditional expectation of X given Y and finally the unconditional expectation of X may be evaluated by taking the expec-

[4] A description of this method is included at the suggestion of a referee.

tations of the resulting expressions with respect to Z and then with respect to Y.

The special equations (1.5), (3.5) (3.6), and (4.5) are derived by using the two following facts:

First,

$$\frac{p}{w+1}\left[D_q + \frac{w}{p}\right] A(n, p, w) = A(n, p, w+1). \tag{7.3}$$

Where D_q means partial differentiation with respect to q. This may be demonstrated by direct application of the operation indicated on the left.

Second, a solution of the difference equation

$$F(w+1, p) = \frac{p}{w+1}\left[D_q + \frac{w}{p}\right] F(w, p) \tag{7.4}$$

(where $q=1-p$)

$$F(w, p) = \frac{p^w}{w!} D_q{}^w F(0, p). \tag{7.5}$$

This may be proved by induction on w, (7.5) being identically true for $w=0$.

For example, (3.5) is derived as follows. From (7.1), putting $n=j-i$

$$P_3\{X = a\} = \sum_n \sum_j A[(j-n)r, s, a-nr] A(j, g, n) A(k, p, j) \tag{7.6}$$

which we write

$$\equiv \sum_n \overline{P}_n(a).$$

$$\overline{P}_n(nr) = \sum_{j=n}^{k} A(k, p, j) A(j, g, n) t^{(j-n)r} \tag{7.7}$$

$$= C_n{}^k g^n p^n (q + pft^r)^{k-n}.$$

Let $w=a-nr$; then $\overline{P}_n$ plays the role of F of (7.5).

Hence,

$$\overline{P}_n(a) = C_n{}^k g^n p^n [(s^{a-nr})/(a-nr)!] D_t{}^{a-nr} (q + pft^r)^{k-n}, \tag{7.8}$$

and

$$(7.9)\quad P_3\{X = a\} = \sum_{n=0}^{k} C_n^k g^n p^n [s^{a-nr}/(a - nr)!] D_s^{a-nr}(q + pft^r)^{k-n}$$

which is equation (3.5) of Table 1.

The same general method is used to get (1.5), (3.6), and (4.5). No equation for large a, i.e., a near kr, is needed in Model 4 because when a is large there are few terms in the sum over i of equation (4.4).

The moments for any model may be obtained from $M_m(\phi)$ by substituting e_ϕ for ϕ, which gives the moment generating function. The αth moment about 0 is then the αth derivative of the moment generating function evaluated at $\phi = 0$.

Alternatively the moments may be evaluated using conditional moments and successively removing the conditions. See, for example, M. H. Hansen, W. N. Hurwitz, W. G. Madow, *Sample Survey Methods and Theory*, Volume II, New York, John Wiley and Sons, 1953, p. 59 ff.

8. EXAMPLE

A numerical example will be worked in detail to show the computing schemes used and to indicate how decisions may be made based on these models. The values used for probabilities and costs in this example are not authentic because true values based on actual experience are classified.

Model 3 is the model assumed for the example. According to this model the probability of an aircraft ditching is p; the probability of getting in the vicinity of the r items which were carried in the ditched aircraft is f; and having arrived in the vicinity of the items the probability of finding a single item is t.

8.1. *Numerical work for the example*

For the example the following numerical values are used:

> kr, the total number of valuable items to be carried, 6. Each of the four possible loadings per aircraft: 1, 2, 3, and 6 is investigated.
> $f = .20$ and $t = .40$ when no salvage device is used;
> $f = .40$ and $t = .60$ with a salvage device. The salvage device might be a radio transmitter which sends automatically when the items are jettisoned.

The expected number of items lost (i.e., ditched and not recovered) is, for any loading:

without salvage device: $6(0.1)\,[1 - (0.2)\,(0.4)] = .55.$
with salvage device: $6(0.1)\,[1 - (0.4)\,(0.6)] = .46.$

The salvage device reduces the expected number of items lost by 0.09 items. If the items are worth a million dollars each, a salvage device which costs less than \$90,000 would be worth developing. The cost must be net cost for salvage devices for all of the six items, less the value of the remaining salvage devices after the ferrying operation has been completed.

The calculation of the probability of losing 0, 1, $\cdots$, 6 items for each of the four possible loadings with and without a salvage device can be arranged so that the computing labor using equation (3.4) is not excessive. The computing scheme shown in Table 2 has been found quite satisfactory.

This scheme is simply a method for systematizing the calculation of the double summation which gives the probability of losing exactly a items, $P\{a\}$. Usually when making decisions based on $P\{a\}$, it is necessary to calculate $P\{a\}$ for all (k, r) sets, and when this is necessary the general formula m.4 for $P\{a\}$ is frequently more efficient than the specialized formulas for $P\{a\}$ which are also given for some models.

Table 2 is for $k=6$, $r=1$, no salvage device. $k=6$, $r=1$ was chosen for the detailed description of Table 2 because this (k, r) pair requires the most extensive computing sheet.

On an actual computing sheet only the numerical values are entered because the literal values are not necessary and may be confusing to the calculator. Blanks in the body of Table 2 are to be read as zeros.

The entries on the computing sheet which are in Roman type are original inputs, based on the assumed values of k, r, p, f, and t. The values in italics and bold face are derived values. The operation being performed in arriving at the italicized values is the multiplication of $P\{a|i, j\}$ by $P\{i|j\}$ for a fixed j and adding over all possible i values. In terms of the formula given in Table 1 the italicized values are:

$$\sum_{i=j-b}^{b} A(j, f, i)A(ir, t, jr - a),$$

where $A(0, t, 0)=1$ by definition. For example, the italicized value of .7787 found in the column for 3 items lost is the sum of products $(.512)(1.00) + (.384)(.600) + (.096)(.36) + (.008)(.216) = .7787$, or in literal notation

$$A(3, f, 0)A(0, t, 0) + A(3, f, 1)A(1, t, 0) + A(3, f, 2)A(2, t, 0) + A(3, f, 3)A(3, t, 0).$$

To find $P\{a\}$ the italicized values are multiplied by the $P\{j\}$ on

TABLE 2

SAMPLE COMPUTING SHEET

No Salvage Device. $k=6,\ r=1,\ p=.10,\ q=.90,\ f=.20,$
$g=.80,\ t=.40,\ s=.60$

Number of aircraft ditched (j)	$P(j)$	Number of aircraft found (i)	$P[i\|j]$	Number of Items Lost: 0	1	2	3	4	5	6
0	A(6, p, 0) .5314	0	A(0, f, 0) 1.00	A(0, t, 0) *1.00*						
1	A(6, p, 1) .3543	0 1	A(1, f, 0) .80 A(1, f, 1) .20	A(1, t, 1) .40 *.08*	A(0, t, 0) 1.00 A(1, t, 0) .60 *.92*					
2	A(6, p, 2) .0984	0 1 2	A(2, f, 0) .64 A(2, f, 1) .32 A(2, f, 2) .04	A(2, t, 2) .16 *.0064*	A(1, t, 1) .40 A(2, t, 1) .48 *.1472*	A(0, t, 0) 1.00 A(1, t, 0) .60 A(2, t, 0) .36 *.8464*				
3	A(6, p, 3) .0146	0 1 2 3	A(3, f, 0) .512 A(3, f, 1) .384 A(3, f, 2) .096 A(3, f, 3) .008	A(3, t, 3) .064 *.0005*	A(2, t, 2) .16 A(3, t, 2) .288 *.0177*	A(1, t, 1) .40 A(2, t, 1) .48 A(3, t, 1) .432 *.2031*	A(0, t, 0) 1.00 A(1, t, 0) .60 A(2, t, 0) .36 A(3, t, 0) .216 *.7787*			
4	A(6, p, 4) .0012	0 1 2 3 4	A(4, f, 0) .4096 A(4, f, 1) .4096 A(4, f, 2) .1536 A(4, f, 3) .0256 A(4, f, 4) .0016	A(4, t, 4) .0256 *.0004*	A(3, t, 3) .064 A(4, t, 3) .1536 *.0019*	A(2, t, 2) .16 A(3, t, 2) .288 A(4, t, 2) .3456 *.0325*	A(1, t, 1) .40 A(2, t, 1) .48 A(3, t, 1) .432 A(4, t, 1) .3456 *.2492*	A(0, t, 0) 1.00 A(1, t, 0) .60 A(2, t, 0) .36 A(3, t, 0) .216 A(4, t, 0) .1296 *.7164*		
5	A(6, p, 5) .00005*	0 1 2 3 4 5	A(5, f, 0) .3277 A(5, f, 1) .4096 A(5, f, 2) .2048 A(5, f, 3) .0512 A(5, f, 4) .0064 A(5, f, 5) .0003	A(5, t, 5) .0102	A(4, t, 4) .0256 A(5, t, 4) .0768	A(3, t, 3) .064 A(4, t, 3) .1536 A(5, t, 3) .2304	A(2, t, 2) .16 A(3, t, 2) .288 A(4, t, 2) .3456 A(5, t, 2) .3456	A(1, t, 1) .40 A(2, t, 1) .48 A(3, t, 1) .432 A(4, t, 1) .3456 A(5, t, 1) .2592	A(0, t, 0) 1.00 A(1, t, 0) .60 A(2, t, 0) .36 A(3, t, 0) .216 A(4, t, 0) .1296 A(5, t, 0) .0778	
6	A(6, p, 6) .000001*	0 1 2 3 4 5 6	A(6, f, 0) .2621 A(6, f, 1) .3932 A(6, f, 2) .2458 A(6, f, 3) .0819 A(6, f, 4) .0154 A(6, f, 5) .0015 A(6, f, 6) .0001	A(6, t, 6) .0041	A(5, t, 5) .0102 A(6, t, 5) .0369	A(4, t, 4) .0256 A(5, t, 4) .0768 A(6, t, 4) .1382	A(3, t, 3) .064 A(4, t, 3) .1536 A(5, t, 3) .2304 A(6, t, 3) .2765	A(2, t, 2) .16 A(3, t, 2) .288 A(4, t, 2) .3456 A(5, t, 2) .3456 A(6, t, 2) .3110	A(1, t, 1) .40 A(2, t, 1) .48 A(3, t, 1) .432 A(4, t, 1) .3456 A(5, t, 1) .2592 A(6, t, 1) .1866	A(0, t, 0) 1.00 A(1, t, 0) .60 A(2, t, 0) .36 A(3, t, 0) .216 A(4, t, 0) .1296 A(5, t, 0) .0778 A(6, t, 0) .0467
				.5604	.3407	.0863	.0117	.0009	.0000	.0000

* The probability of 5 or 6 aircraft ditching is so small that it is unnecessary to carry out the calculations for these values, and all entries after those for $j=4$ would be omitted on an actual computing sheet. The entries are given in Table 2 to show in detail how the complete computing sheet is set up.

the same line and these products are summed for each column. The results are the bold face values at the bottom of the columns, which are the $P\{a\}$ values. From Table 2:

$$P\{0\} = .5604 = (.5314)(1.00) + (.3543)(.08) + (.0984)(.0064) + (.0146)(.0005) + (.0012)(.0004)$$

$$P\{1\} = .3407$$

$$\cdots\cdots\cdots$$

$$P\{4\} = .0009$$

$$P\{5\} = P\{6\} = .0000.$$

Table 3 summarizes the probability values for the example. Values are presented for the four possible (k, r) pairs, each with and without a salvage device. The individual probabilities are given in the left hand column of each pair of columns, and cumulated values of the probabilities are given in the other column. The cumulated values show the probability of losing at least so many items. For example, in Table 3, the .440 in the third column means that the probability of losing one or more items (at least one item) is .440.

8.2. Decision making for the example[5]

If one loading resulted in a probability of loss which was less than or equal to that of any other loading for all number of items lost, that loading would be uniformly best.

The necessity for executive decision is based on the fact that this is not so. If one wishes to minimize the probability of losing one or more items, the best loading is $k=1$, $r=6$. But if one wishes to minimize the probability of losing 3 or more items, the best loading is $k=6$, $r=1$. However, the loading $k=3$, $r=2$ yields almost as low a probability of losing 3 or more items, as the loading $k=6$, $r=1$, and the cost of running the operation has been halved since only half as many aircraft are involved in the operation. It appears from Table 3 that for the parameter values involved in the example the loading is a more important variable than the use or non-use of a salvage device.

One might wish to minimize the probability of losing 6 items if the crucial consideration were to deliver at least one item. This would be true for example if the items were special code books. One code book might be absolutely essential for the performance of an operation and

[5] A referee suggests that reference be made to the recent book by D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, New York, John Wiley and Sons, 1954, where a detailed discussion of statistical decision theory may be found.

TABLE 3

SUMMARY OF PROBABILITIES FOR THE EXAMPLE

No. of items lost	($k=6$, $r=1$)				($k=3$, $r=2$)				($k=2$, $r=3$)				($k=1$, $r=6$)			
	No salv. device		With salv. device		No salv. device		With salv. device		No salv. device		With salv. device		No salv. device		With salv. device	
	Indiv. terms	Cum.*	Indiv. terms	Cum.*	Indiv. terms	Cum.*	Indiv. terms	Cum.*	Indiv. terms	Cum.*	Indiv. terms	Cum.*	Indiv. terms	Cum.*	Indiv. terms	Cum.*
0	.560	1.000	.622	1.000	.737	1.000	.765	1.000	.812	1.000	.826	1.000	.900	1.000	.902	1.000
1	.341	.440	.307	.378	.023	.263	.048	.235	.010	.188	.031	.174	.001	.100	.007	.098
2	.086	.099	.063	.071	.214	.240	.168	.187	.016	.178	.021	.143	.003	.099	.012	.091
3	.012	.013	.007	.008	.004	.026	.007	.019	.152	.162	.114	.122	.005	.096	.011	.079
4	.001	.001	.002	.001	.021	.022	.012	.012	.001	.010	.002	.008	.006	.091	.006	.068
5	.000	.000	.000	.000	.000	.001	.000	.000	.002	.009	.002	.006	.004	.085	.002	.062
6	.000	.000	.000	.000	.001	.001	.000	.000	.007	.007	.004	.004	.081	.081	.060	.060

* The entries in this column are the sums of the values in the preceding column cumulated from the large values of number of items lost.

the delivery of all six code books rather than just one would simply increase the ease with which the operation could be performed. Under these conditions one would ordinarily choose the loading that minimized the probability of losing 6 items, even if this loading led to the highest probabilities for all other numbers of items lost. This is an example of a rather typical situation: it is impossible to put a monetary value on the items being carried, and hence executive judgment is essential.

If the items were key executives of a corporation, it might be considered essential to have four of them survive. In this case one would wish to minimize the probability of losing 2 or more items (executives). From Table 3 it appears that $k=6$, $r=1$ or $k=1$, $r=6$ are almost equally good. If in this case one had no salvage device $k=1$, $r=6$ would be the choice since this costs less than $k=6$, $r=1$ and the probabilities are equal. If one has a salvage device, however, $k=6$, $r=1$ is a better loading than $k=1$, $r=6$, since it leads to a slightly lower probability of losing 2 or more items (.071 *vs.* .091). The decision as to which loading to use depends on the extra cost of making six trips instead of one compared to the extra utility resulting from a reduction of 0.02 in the probability of losing 2 or more executives. But here again the utility cannot be measured in money and the decision must be made by executive judgment.

THE EXPERIMENTAL APPROACH IN THE TEACHING OF STATISTICS*

Edwin G. Olds
Carnegie Institute of Technology

1. INTRODUCTION

At a session of the American Statistical Association, held five years ago at Cleveland, I presented a paper[1] on the use of instructional aids in teaching statistical quality control. At that time it was noted [2, pp. 223–24] that, in the past, the average teacher of elementary probability had made little use of experiments in his teaching and the remark was made that

> It is hard to understand why he failed to appreciate the pedagogical value of designing an experiment to illustrate a point of theory, predicting the result, running the experiment, and then taking the consequences if it turned out wrong.

Clearly, the existence of some "value" in the use of experiments in teaching statistics was implied.

In view of the time which has elapsed, it seems appropriate to take a fresh look at the matter.

2. TYPES OF EXPERIMENTS

In a recent dictionary [4, p. 352] an experiment is defined as:

> A trial made to confirm or disprove something doubtful; an operation undertaken to discover some unknown principle or effect, or to test some suggested truth, or to demonstrate some known truth; . . .

The four classes of trials can be reduced to two by combining aims, one, three and four. Then the types of experiments described in the definition might be classified as those performed:

(1) to test hypotheses,
(2) for exploration.

3. DEMONSTRATION LECTURES

It would be possible to cross-classify experimentation in learning statistics on the basis of who performs the experiment. The so-called experiments used in demonstration lectures may be done by the

* Presented under a slightly different title at the Annual Meeting of the American Statistical Association at Washington, D. C., December 29, 1953.

[1] This paper appeared as Part III of a joint publication [2]. Bracketed numbers refer to the list of references at the end of this paper.

teacher, by the students, or both both working together. Any individual student may be a participant or an observer.

Upon reviewing my own demonstrations, it is not at all clear how many of them should be classified as experiments at all. For analysis, a lecture-plan from a manual [5], prepared for use in the intensive courses in statistical quality control given in various industrial centers during World War II, is presented below. Some description of the equipment used is given in [2, pp. 224–25] but it is sufficient to note that the bead-box of reference ordinarily contained 1152 white beads and 48 red beads.

II, 4.—THE CONTROL CHARTS FOR FRACTION DEFECTIVE AND NUMBER OF DEFECTIVE ITEMS

Objective

1. To present the uses and purposes of a chart for fraction defective or number of defective items.
2. To give the techniques necessary for the construction and utilization of these charts.
3. To demonstrate the variation of fraction defective in samples drawn from a lot or process having a constant proportion of defective items.
4. To indicate the sensitivity of the limits in detecting a process change.

Procedure

1. Make introductory remarks on the uses of one of these charts, including the fact that the necessary data may be readily available in the form of day-by-day accounting records on one hundred per cent inspection.
2. Point out that the box of beads is to represent a lot of material produced by a machine; that, after each sample, we must imagine that the lot is removed and a new lot presented. This means, of course, that the lots are uniform and the process is controlled. The samples should verify this fact.
3. Take three or four samples of 50 beads in order to indicate the procedure and method of computation.
4. Draw and record twenty samples.
5. Plot values for number of defective items and for fraction defective.
6. Compute values for central line and control limits, explaining that the limits are 3-sigma limits.
7. Put limits on both charts, then point out that they tell the same story. At this point discontinue consideration of chart for number of defective items.
8. Examine chart for indication of control and state that since we have analyzed past data and, seem to be in control, we can use $\bar{p}$ as the standard value, p', and extend the control limits for use during production.
9. Have students make charts for use in controlling the process and prepare to record and plot new sampling results.

10. Increase the percentage of defectives in the box and draw samples until a point falls outside the control band.
11. Tell the students about the process change at the beginning of the second set of drawings, pointing out that it did not show itself immediately.
12. Continue sampling until there are twenty samples in the second set.
13. Have the students compute the limits for the new set of drawings and place them on the chart.
14. Give students the value of original fraction defective in the bowl; point out that, since the first twenty samples of 50 were in control, their total could be considered as a single sample of 1000.
15. Using original fraction defective, calculate control limits for samples of 1000 and point out relationship of these limits to the first and second values of $\bar{p}$.
16. Return to consideration of the chart having two sets of limits and discuss the significance of the area common to the two bands.

Principal points for emphasis

1. Charts of this kind have proven to be very useful and easy to explain to management.
2. Many students have found it desirable to start their use of control charts by analyzing data on fraction defective.
3. Unless samples are quite large, these charts are not very sensitive to small changes.
4. Control charts for measurements are, in general, to be preferred. They give much more information for process analysis.

A critical review of the outlined procedure raises the question as to what the student learns from it. Unless ample time is used to give him careful explanation he may learn little. I am convinced that much experimental work falls short in garnering full educational value because the lecturer has been niggardly in the time allotted to orientation.

Given sufficient time and some imagination an instructor can teach a good proportion of an elementary statistics course from just one demonstration of this type.

In the first place, it can be pointed out that the bead box represents a universe or population. The population is finite and well-defined. Other populations of interest to statisticians may not be so well-defined. They may be so large as to seem almost infinite.

Secondly, the individual units can be readily classified into two groups on the basis of color. If a red bead is called a defective, there will be little argument as to whether or not an individual unit is defective

Third, the universe is characterized by a single parameter, denoted by p'. If weight were the quality under consideration, the classification would be based on measurements and the finite universe would be

described by a frequency tabulation, needing at least two parameters to summarize it adequately.

Fourth, if we are interested in finding out what proportion of the population is defective, we might count the beads. In general, this procedure is too expensive so we must depend on the uncertain evidence from a sample. Obviously a single unit is not sufficient to represent the universe so we need to examine several units. How large a sample should we take? How should we take it? What statistic should we calculate?

A discussion of the above questions can carry us through the elements of probability; the binomial and hypergeometric distributions, together with the Poisson and normal distributions as approximations to them; an introduction to point and interval estimation; tests of hypotheses; and acceptance sampling. Furthermore, the need for planned experimentation becomes evident.

It might be kept in mind that all of this can be discussed before beginning the scheduled demonstration. With the box of beads and the sampling paddle as actors we have given reality to the characters of our statistical play. Our audience is well acquainted with their human frailties and will not be too much surprised if they misbehave.

It does not seem necessary to fill in the details for the entire demonstration, or to call attention to the opportunities for problem and theory assignments based on it. It might be remarked, however, that one performance of Steps 4–7 of the indicated procedure provides a single sample of the operation of the control chart for fraction defective. Assuming that the data has been produced by a controlled process of unknown fraction defective, this sample of one can be used to estimate the probability that the recommended procedure will err in indicating lack of control. (Step 8 is rather interesting because it does not seem to recognize the possibility of such an error.) Alternatively, the sample of one can be used to test some hypothesis regarding the probability of this error of the first kind.

The later parts of the indicated procedure provide other single samples of control-chart operation under various conditions. These samples could serve the purposes indicated above or they could be combined with the first sample to provide a sample of one from the population of all repetitions of the complete procedure as outlined. Parenthetically, the opportunity here for confusion by the student concerning sample size and the population sampled should be noted.

Granted that, as a demonstration, the quoted procedure may have merit, the question remains as to whether or not it should be called

an experiment. The answer seems to depend on the attitude of the instructor and students toward it. As indicated above, it is possible to plan and execute the demonstration as one or several experiments. Furthermore, each may be performed for exploration or to test some hypothesis.

4. LABORATORY WORK

It is my impression that much of the laboratory work in statistics consists of working problems involving a considerable amount of calculation. Practice in the choice and use of formulas is valuable and there can be no quarrel with the usefulness of learning how to operate calculating machines. The question arises, however, whether it would be possible to broaden laboratory work to include a few experiments.

In a paper [3] presented at the Chicago meeting in December, 1952, Professor A. C. Rosander proposed a list of some forty-five laboratory experiments for probability statistics and suggested that these be organized into a manual. The preparation of such a manual would seem to be a useful enterprise. It would seem difficult, however, to avoid loading the manual with experiments requiring an undue amount of mechanical repetition and uninteresting computations.

Industrial laboratories are beginning to be interested in statistically planned experimentation. For this work the cooperation of statisticians skilled in experimental design will be sought. It would seem desirable to begin the training in design by work in the statistical laboratory. Therefore, a manual such as Rosander proposes ought to include exercises on designing experiments as well as on executing them.

5. THE EXPERIMENTAL APPROACH IN THE SEARCH FOR TRUTH

Up to this point I have considered the use of the experiment as an aid in digesting the existing body of statistical principles and methods. There is quite another aspect of the experimental approach which now deserves some attention. I refer to the use of experimentation in pushing back statistical frontiers.

The rôle of induction and generalization in the field of statistics is well known, with the contribution of "Student" as a shining example. It seems reasonable to expect that the study of samples will continue to feed the intuition needed to bridge the gap between the experimenter and his goal.

With the greater availability of high-speed computers experimental sampling can be done on a scale undreamed of fifty years ago. A

Weldon[2] who rolls a set of 12 dice a total of 26,306 times yields his place in print to the machine which turns out a few million pseudo-random numbers while warming up for a really big job. The inductive reasoner no longer need restrict his operation to a few small samples.

As a footnote to this discussion of using the experimental approach in the search for truth, an additional remark might be added. Belief is an individual matter. Some students require one type of proof, some another. One of my beginners could not accept the mathematical derivation of the probability density function for the sum of a sample of two from a rectangular universe. After he and his wife spent several hours drawing samples from a double deck of cards he was entirely satisfied with the truth of the theorem.

This is, by no means, an isolated example. It is my belief that a considerable proportion of our elementary students have little faith (or interest) in mathematically established truths until they have seen experimental verification. One definition of "experiment" quoted above was

"A trial made . . . to demonstrate some known truth"

For the thousands of men taught statistical quality control in the past dozen years the trials made to demonstrate the instructor-known truths often have seemed to be the only means of getting the message through to its destination.

6. EVALUATION OF RESULTS

My previous remarks regarding the experimental approach imply that it has some positive value. This opinion might well be tested by means of a designed experiment. Presumably any scientifically-minded educator who advocates a particular plan of teaching is willing to give his pet theory an objective test. Statisticians, in particular, claim that they prefer to base conclusions on facts rather than opinions.

For our statistician-seedlings we might have two sets of treatments. The first set, as discussed in Section 2 above, might consist of the exploratory experiment and the experiment to test hypotheses. The second set of treatments might include the experiment by the instructor in lecture (with or without help) and the experiment by the student in the laboratory.

The treatments in the second set could be applied in various strengths. Either we could rush through the experiment, or we could provide various amounts of orientation before, during and after per-

[2] Weldon's dice data are given in [1, p. 278].

formance. I have averred that the student learns little unless the ground-work is carefully laid; but this is an opinion needing proof.

Measurements of two kinds of results should be made on our subjects. In the first place, increase in knowledge of existing principles and methods should be gauged. Second, there should be a measurement of discovery-value.

Finally, the planned procedure ought to provide for a control group which receives no teaching by the experimental approach.

It is not my intention to extend the discussion of this test any further beyond remarking that in order to plan and conduct it properly the cooperation of a professional in the field of educational tests and measurements probably would be necessary.

7. CONCLUSION

In the early part of this paper the opinion has been expressed that the experimental approach has value and a few suggestions have been made regarding combinations of treatments for best results. In the last section it has been suggested that these hypotheses should be tested by a planned experiment. Whether or not there is agreement with the earlier statements in this paper, I am confident that the recommendation for the collection of some unbiased evidence will receive support.

REFERENCES

[1] Fry, T. C., *Probability and Its Engineering Uses*, New York, Van Nostrand (1928).

[2] Olds, Edwin G., and Knowler, Lloyd A., "Teaching statistical quality control for town and gown," *Journal of the American Statistical Association*, 44 (1949), 213–30.

[3] Rosander, A. C., "The use of laboratory experiments in teaching probability statistics" (unpublished manuscript).

[4] *Webster's Collegiate Dictionary, Fifth Edition*, Springfield, Mass., Merriam (1947).

[5] Working, H., and Olds, E. G., *An Introduction to Statistical Methods of Quality Control in Industry: An Outline of a Course of Lectures and Exercises*. Quality Control Program, Carnegie Institute of Technology (1945). Prepared for Office of Production Research and Development, War Production Board, and now available at Government Printing Office, Washington, D. C.

USE OF EXPERIMENTS IN TEACHING ENGINEERING STATISTICS*

IRVING W. BURR
Purdue University

1. INTRODUCTION

EXTENSIVE use of sampling experiments was one of the characteristic features of the war-time intensive courses in statistical quality control [5], and is still an integral part of the present short courses in the subject. For those with as little mathematical skill and understanding as many industrial men have, derivations are out of the question, and the only recourse is to demonstrate the theory by sampling experiments. Industrial men are readily convinced by such experiments, especially when they themselves do the sampling.

The situation would seem to be quite different when one is teaching engineering students who have had calculus. It is not as different, however, as it seems. Engineering students are intensely practical and few of them are fond of mathematical theory. Hence well chosen and carefully explained experiments are a valuable supplement to mathematical derivation. Even for a student majoring in mathematics, experiments can serve to illuminate the theory, and they certainly do give the average student a clearer idea of "statistical thinking."

The instructor should make clear (*a*) the principles involved, (*b*) the necessity for comparison of the observed results with theory, (*c*) the careful arrangement for randomness, and (*d*) the fallibility of the experiment, explaining any "misbehaving" of the experiment.

2. SAMPLING POPULATIONS

For fraction defective experiments a convenient way to simulate a binomial distribution is to use a box of beads.[1] One color, say white, can be called good, and another, say red, can be called defective. Then if we use a great many more beads in the box than in a sample, the actual hypergeometric distribution will give a good approximation to the desired binomial. It is convenient to maintain 1000 white beads and then vary the number of red ones for various fractions—42 for 4 per cent (42/1042), 87 for 8 per cent, etc. Samples of 50 seem to work well. A description of paddles for such drawing is given in part III of

* Presented at the Annual Meeting of the American Statistical Association at Washington, D. C., December 29, 1953.

[1] Suitable beads one centimeter in diameter of various colors may be obtained from Walco Bead Co., 37 West 37th St., New York City.

an article by Olds and Knowler [4]. The *relative* error in approximating a binomial probability by a hypergeometric probability is approximately [2]

$$-\frac{1}{2Np'q'}\left[d-(d-np')^2\right],$$

where N and n are the lot and sample sizes respectively, d is the num-

TABLE 1

POISSON POPULATIONS FOR SAMPLING EXPERIMENTS

No. of Defects	Number of Chips for Given Parameter				
	$c'=1$	$c'=2$	$c'=4$	$c'=6$	$c'=8$
0	184	68	9	1	0
1	184	135	37	7	1
2	92	135	73	22	5
3	31	90	98	45	14
4	8	45	98	67	29
5	2	18	78	80	46
6		6	52	80	61
7		2	30	69	70
8		1	15	52	70
9			7	34	62
10			3	21	50
11			1	11	36
12				6	24
13				3	15
14				1	8
15				1	5
16					2
17					1
18					1
	501	500	501	500	500
Actual Average	1.004	2.006	4.002	6.020	8.012
Actual Variance	1.006	2.042	3.982	6.036	7.916

ber of defectives in the sample, p' is the fraction defective in the lot, and q' is $1-p'$.

For the Poisson distribution, no such convenient method is available, other than that of making the double approximation of the hypergeometric for the Poisson. The best possibility would seem to be to

use beads or chips, each numbered with a number of defects.[2] Then drawing one chip or bead gives a value of c, the number of defects. For such a Poisson population it is desirable to have at least 500 pieces so that there will be a few rare values of c available in the population, in order that a c chart can go out of control. If too few chips, say 100, are in the population there probably will be no rare values of c—none can be rarer than 1 in 100. The populations shown in Table 1 are quite serviceable. It should be noted that if a sample from a population with a larger parameter, c', is desired, this is readily available because of the additive property of the Poisson distribution. Thus, if we want $c'=14$, we can draw one chip from $c'=8$ and one from $c'=6$, and the total of the two c values will have the desired distribution.

For normally distributed populations Table 2 gives a flexible set of distributions. These were used in the war-time courses and are still in wide use. Another convenient way to generate approximately normal populations is through using various numbers of dice. Although the distribution of the number of points on a single die is rectangular, the total number of points on 3 dice is fairly normal, as the following theoretical distribution shows:

Total 3 dice	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
216 × Probability	1	3	6	10	15	21	25	27	27	25	21	15	10	6	3	1	216

For n dice at a throw we have the following theoretical results for the total points:

$$\overline{X}' = 3.5n, \qquad \sigma' = \sqrt{\frac{35n}{12}}, \qquad \alpha_4 = 3 - \frac{1.27}{n}.$$

3. SAMPLING EXPERIMENTS

In any one course one would never use all of the experiments here suggested. Nevertheless it seems desirable to list them all.

3.1. *Sample Frequency Distribution*

A simple and effective demonstration of the binomial may be had by casting 6 dice and counting the number of aces appearing on each such cast. One hundred casts will give an interesting comparison between the observed frequencies of numbers of aces and the theoretical, from $100\left(\frac{1}{6}+\frac{5}{6}\right)^6$. The chi-square test of goodness of fit may be used, if desired.

[2] White fiber chips may be bought from Lamb Seal and Stencil Co., 824 13th St., N.W., Washington, D. C. However, if one can find the right industrial company he can obtain thousands of fiber di[illegible], since they are scrap in many processes.

Binomial samples from beads may be taken with much smaller fraction defective, thus illustrating better the typical situation in industrial lots. Samples from any of the populations in Tables 1 and 2, or dice totals make good examples to compare theory and sample.

TABLE 2

APPROXIMATELY NORMAL POPULATIONS

No.	Frequencies for Population				
	A	*B*	*C*	*D*	*E*
−10				1	
9				1	1
8				1	1
7		1		3	1
6		3		5	3
5	1	10		8	5
4	3	23		12	8
3	10	39	1	16	12
2	23	48	3	20	16
−1	39	39	10	22	20
0	48	23	23	23	22
+1	39	10	39	22	23
2	23	3	48	20	22
3	10	1	39	16	20
4	3		23	12	16
5	1		10	8	12
6			3	5	8
7			1	3	5
8				1	3
9				1	1
10				1	1
+11					1
Total	200	200	200	201	201
Mean	0	−2	+2	0	+1
Std. Dev.	1.71	1.71	1.71	3.47	3.47
α_4	3.02	3.02	3.02	3.02	3.02

3.2. Control Charts

Effective use of the foregoing populations may be made to illustrate $\bar{X}$, R, σ, p, np, and c charts. One may take samples to compare against center line and limits calculated from the known population characteristics. But perhaps more interesting is the case of "no standard

given." A series of preliminary samples are drawn and control lines figured. Control is checked, the lines continued, and new "production" analyzed. The population can then be changed by increasing p' for fraction defective, or by using a different measurement or Poisson population. Thus one might use Population A of Table 2, then shift to B or D, or use $c'=4$ in Table 1, then shift to $c'=8$ or 1, and observe the effect on the chart.

Other interesting experiments have to do with ways in which samples can be taken [5], [2], [3]. For example, a stratified sample may be taken by letting X_1 be the total points for 2 dice, X_2 that for 3 dice, X_3 for 4 dice, X_4 for 5 dice. Then the first X_1, X_2, X_3, and X_4 together comprise the first sample, etc. Such a sample is stratified because it contains one value from each of 4 populations. The range for such samples will be so inflated by the difference between means, especially X_1 and X_4, that both charts for $\bar{X}$ and R will show "too good" control. On the other hand if the first four X_1's from the same data are taken in one sample, then four X_2's, etc., and all the data put on one $\bar{X}$ and R chart the $\bar{X}_1$'s and $\bar{X}_4$'s will stand out beyond the control limits like "sore thumbs," because they are out of control with respect to the rational sample variability. The same experiment may be readily done with 3 distributions A, one of B and one of C. Such populations can then be mixed, and samples drawn, illustrating how random samples from mixed product show good control despite the presence of assignable causes.

3.3. Significance Tests

Any of the standard significance tests may be made by using a sample from one of the populations. Thus one can test the hypothesis that $\bar{X}'=0$, using either population A or D, and assuming either that σ' is known or unknown. Or one can test the same hypothesis against populations B, C, or E. An interesting variation is to let each class member make the test by drawing his own sample. In this way one would expect about one class member in 20 to refute a true hypothesis, at the 5 per cent level. Also the class can thus build up a t, or a non-central t distribution.

3.4. Estimation

The subject of biased and unbiased estimates of population parameters may be illustrated by drawing a series of small samples and tabulating the various kinds of estimates, say, of both population standard deviation and variance.

Confidence intervals may be set from each of a series of samples. There is a beautiful illustration of this kind of experiment given by Shewhart [1]. It is advisable to use 90 per cent confidence intervals so that there is a reasonable chance for an interval to fail to contain the parameter.

3.5. *Analysis of Variance*

Obviously one can run an analysis of variance with all cell samples from distribution A, to illustrate the null hypothesis. Data for simple designs can be repeatedly drawn and calculations made to yield a distribution of F values. Then one or two cell samples can be drawn from B or C and thus the null hypothesis stands to be rejected some of the time at least. (Of course one can more easily illustrate the formation of an F distribution with pairs of samples from A, or say from 4 dice, and the non-central F by a sample from A and one from D, or one with 4 dice then one with, say, 10 dice.)

Less obvious but equally valuable for illustration are experiments where the true cell mean is determined from any desired linear hypothesis, and then a random error component is added, such as a drawing from distribution A, or a toss of 3 dice. A linear or quadratic trend can readily be simulated. Such trends must be strong relative to the random error, unless large samples are to be drawn at each point.

Tests for homogeneity of variance, such as Bartlett's, and homogeneity of proportions, such as chi-square, also can be readily illustrated.

3.6. *Statistics of Combinations*

One good experiment to illustrate the tolerances for mating parts is often worth any amount of equations for the practical man. A successful one is the following:

Outside diameter of shaft

$$\bar{X}' = 3.00105'', \qquad \sigma' = .000296''.$$

Throw 3 dice, and regard the number of points as the number of ten-thousandths of an inch in excess of three inches.

Inside diameter of bearing

$$\bar{X}' = 3.0021'', \qquad \sigma' = .000418''.$$

Throw 6 dice, and regard the number of points as the number of ten-thousandths of an inch in excess of three inches.

Draw the two frequency distributions, the 3σ limits for each and the true extreme possible values for each. The percentage of interference is only about 2.6 per cent as may be illustrated by experiment, but guesses prior to experiment, based on the two frequency distributions shown, will usually run much higher. Those who want *no overlapping* in even the extreme ranges, would never sanction such distributions, even though they would probably be satisfactory in practice. Hence such persons would specify a considerably greater mean difference, thus giving too many loose fits.

3.7. *Acceptance Sampling*

A wide variety of sampling experiments are regularly used to show the way various sampling inspection plans operate on "acceptable," "marginal," and "rejectable" material, or on mixed product. It can be shown how the consumer is protected against bad quality, on a lot-by-lot basis and on an average-quality basis. An empirical operating characteristic curve, showing the observed probability of acceptance for given quality may be built up. Not only attribute plans, but also variables plans may be illustrated in operations. For other possibilities,see the original manual [5] and current manuals for short courses in quality control, such as those at the Universities of Iowa, Illinois, and Michigan.

3.8. *Sequential Analysis*

Experiments are especially effective in showing the way in which sequential analysis reaches its decision. The same plan can be tried on different populations, some "good" and others "bad." Any of those listed in Section 2 can be so used.

3.9. *Linear Correlation*

One can easily illustrate sampling from an uncorrelated population by drawing independent random samples from the pair of populations. For example, let Y be the total points for a throw of 3 dice, while X is a drawing from population A. To illustrate a correlated population, one can draw a chip from A, and then use this as a correction to whatever one gets in a toss of 3 dice. Thus if a "minus 2" is drawn for X, and the throw yields 13, we record 11 for Y, etc. For this case, we have $r = .5013$, and a good approximation to a normal bi-variate population. It can in fact readily be shown that

$$r \doteq \sqrt{\sigma_X^2/(\sigma_X^2 + \sigma_Z^2)}$$

where σ_X^2 is the variance of the X values, and σ_Z^2 the variance of the Y values before given the adjustment from X.

If weaker population correlations are desired, we may throw more dice for the original Y value thus increasing σ_Z^2. Stronger correlations than .5 are obtainable through using distribution D for X and, say, 3 dice for the original Y, or else letting X be a drawing from distribution D, and Y this drawing plus a drawing from A. The respective values of r are .7608 and .8965.

3.10. Curve-fitting

In curve-fitting one can take data from an exact mathematical curve at equally spaced X's, and find one Y value for each such X, by taking the mathematical value and adding a random error. The latter should be small relative to the variation in the mathematical curve if a close relationship is desired. The formula in Section 3.9 can be used to give the true correlation ratio for the non-linear population. Thus we have

$$\rho = \sqrt{\frac{\sigma_W^2}{\sigma_W^2 + \sigma_Z^2}}$$

where σ_W^2 is the variance of the true curve ordinates and σ_Z^2 is the variance of the error component added to W.

3.11. Research Problems

Frequently a research worker is unable to find the distribution of some statistic. Recourse may then be had to samples from a suitable population. For example, an engineer in design wanted the frequency distribution of a compounding of eccentricities at random angle. Thus he wanted the distribution of Z in

$$Z^2 = X^2 + Y^2 + 2XY \cos \theta$$

where the distributions of X and Y were known and θ was rectangularly distributed 0° to 180°. The approximate mean and variance were found, but a sampling experiment was needed to determine the approximate shape of the distribution.

A typical class experiment would be to find empirically the sampling distribution of the standard deviation for a skewed or a J-shaped population.

4. CONCLUSION

The foregoing experiments could form the bulk of the theoretical discussion of a course or could be used only as an occasional supple-

ment to derivations. If carefully explained and interpreted relative to some practical situation, such experiments can be most helpful.

REFERENCES

[1] ASTM *Manual on Quality Control of Materials*, American Society for Testing Materials, Philadelphia, 1951. See p. 45.

[2] Burr, Irving W., *Engineering Statistics and Quality Control*, New York, McGraw-Hill (1953). See p. 209 for correction formula, pp. 274–80 for kinds of sampling.

[3] Burr, Irving W., "Some experiments illustrating principles of quality control," *Quality Control Report 12*, Office of Production Research and Development, War Production Board, Washington, D. C., September, 1945.

[4] Olds, Edwin G., and Knowler, Lloyd A., "Teaching statistical quality control for town and gown," *Journal of the American Statistical Association*, 44 (1949), 213–30.

[5] Working, Holbrook, and Olds, Edwin G., *An Introduction to Statistical Lectures of Quality Control in Industry: An Outline of a Course of Lectures and Exercises*, Office of Production Research and Development, War Production Board, Washington, D. C., April 1944. Presently available at U. S. Government Printing Office.

ERRATA

Readers and authors are invited to submit corrections to papers published in any previous issue. These will be published each year, in the December issue.

Grab, Edwin L., and Savage, I. Richard, TABLES OF THE EXPECTED VALUE OF 1/X FOR POSITIVE BERNOULLI AND POISSON VARIABLES, Vol. 49, No. 265 (*March 1954*), 169–77.

The following paper has recently been drawn to our attention: J. Tiago de Oliveira, "Sur le calcul des moments de la réciproque d'une variable aléatoire positive de Bernoulli et Poisson," *Anais da Faculdade de Ciências do Porto*, 36 (1952), 5–8.

The material in de Oliveira's paper suggests alternative methods for computing the tables of our paper.

Laderman, J., Littauer, S. B., and Tukey, John W., THE INVENTORY PROBLEM, Vol. 48, No. 261 (*December 1953*), 717–732.

The following corrections should be made:

Page	*Line*	*Reads*	*Should Read*
723	12 (2nd display line)	for $d \leqq y$.	for $d \geqq y$.
726	16	henec	hence
729	17	$x \leqq_2 5$	$x_2 \geqq 5$

Roshwalb, Irving, EFFECT OF WEIGHTING BY CARD-DUPLICATION ON EFFICIENCY OF SURVEY RESULTS, Vol. 48, No. 261 (*December 1953*), 773–777.

The first term within the brackets of expressions (3) and (6) should read $4n(P-N)/(P-1)$ instead of $4n(P-n)/(P-1)$. As a consequence, the coefficient of r in the denominator of (4) should be $(1+d)^2$ instead of $(5d^2-2d+1)$. These changes do not affect the tabulated relative information in the case of sampling with replacement and have no significant effect except in the case of high sampling rates.

Savage, L. J., THE THEORY OF STATISTICAL DECISION, Vol. 46, No. 253 (*March 1951*), 55–67.

On page 60, in lines 19, 32, and 36 change \$1 to \$1/2; and in line 36 change \$2 to \$4.

STATISTICAL ABSTRACTS

All communications concerning this section should be addressed to the Abstracts Editor, Professor George E. Nicholson, Jr., Chairman of the Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Ardian, Leo A., "What Makes a Quality Control Chart Tick," *Industrial Quality Control*, 10 (1954), 38–43.

The concept of errors of type I and type II are presented for control charts based on fraction defective and number of defectives. Tables and charts are included showing the relation between the two types of errors as functions of selected sample sizes and quality levels. Two probability models are considered when the process is out of control: (1) it is operating at a new constant level, (2) it is operating at a different level at each decision point. GERALD J. LIEBERMAN, *Stanford University*.

Borch, Karl, "Effects on Demand of Changes in the Distribution of Income," *Econometrica*, 21 (1953), 325–31.

This paper reports the results of a particular attempt to lend a more explicit economic interpretation to effects frequently attributed to "time" variables in econometric analyses. The particular empirical results on which this further analysis is based are those of Prest (*The Review of Economics and Statistics*, 31: 1, 1949), as revised by Farrell (*Econometrica*, 20: 2, 1952), for beer, spirits, and tobacco. Prest analyzed demand for these three goods in the United Kingdom, using time series data for the period 1870–1938, omitting 1915–1919. The basic form of the demand relation used by Prest was, $C_t = kY_t^a P_t^b e^{ct+dt^2+fx}$, where C and Y are, respectively, consumption and income per capita, P is price deflated by a cost-of-living index, t is time in years, and x is a discontinuity variate taking value 0 for 1870–1914 and 1 for 1920–1938. In Prest's results the time variables in the exponent were dominant variables in "explaining" the variation in consumption over the period. The present author's hypothesis is that changing income distribution over time might be an important economic factor underlying the highly significant coefficients of the time variables in Prest's results.

The author presents results of a particular formulation consistent with his hypothesis. The form of the income distribution function is assumed to be the logarithmic-normal and income elasticity of consumption is assumed to be represented by $E(y) = p + q/\log y$. In this framework the pattern of change in income distribution over time which would approximately account for the effect of the trend found by Prest is determined. Taking Farrell's revised estimates of the coefficients a, c, d, and f as given, the coefficients p and q in the income elasticity expression are approximated. Having values for p and q, the coefficient of variation of the income distributions at selected points in time are calculated. The results, apart from extreme values for spirits in two years, reflect a marked trend toward equality in the distribution of income. On *a priori* grounds this is considered an acceptable development of the income distribution. The p and q obtained are also used to calculate income elasticities corresponding to different levels of income. It is concluded that the resulting elasticities do not generally conform to what would be expected intuitively. A main conclusion of the author is the following: "One should not draw overly general conclusions from the rough calculations in this paper. The results seem, however, to indicate that changes in the distribution of income can play an important part in explaining time trends in demand functions." IVAN M. LEE, *University of California*.

Brown, T. M., "Standard Error of Forecast of a Complete Econometric Model," *Econometrica*, 22 (1954), 178–92.

The author develops and presents in matrix form approximate formulas for the estimation of the elements of the vector of the standard error of forecasts of the endogenous variables in a multiequation econometric model. The framework is developed first with respect to a single equation in which it is assumed that the conditions of the Markoff Theorem are met. For the single equation case, a general expression for the forecast variance is developed as a sum of two components, viz., the variance of the estimated mean of Y for given Z

and the disturbance variance. This may be expressed in matrix form as, $\sigma^2(Y_F) = Z_F[\sigma(a)]Z_F' + \sigma^2$, where Z_F is the vector of predetermined variables assumed known) in the forecast period, $\sigma(a)$ is the covariance matrix of estimated coefficients, and σ^2 is the disturbance variance. In the single equation case, $\sigma(a)$ and σ^2 are estimated by well-known methods.

In the multiequation complete econometric model, the above forecast variance becomes a vector containing as many elements as there are endogenous variables in the system. Each element of this vector may be viewed as the sum of two components analogous to those specified for the single equation case. The multiequation complete econometric model may be expressed, $\beta Y' + \Gamma Z' = AX' = \mu_p'$ where Y, Z and μ_p are, respectively, vectors of endogenous variables, predetermined variables, and unobservable disturbance terms; β and Γ are population coefficient matrices; $X = [YZ]$; and $A = [\beta\Gamma]$. Let the estimated model be represented by $BY' + CZ' = AX' = u_p'$, where B, C, and u_p are estimates of β, Γ, and μ_p, respectively. Written in a form most suitable for forecasting Y for given Z, the estimated model becomes, $Y' = B^{-1}CZ' + B^{-1}u_p' = FZ' + u_s \cdot u_p$ have been designated "partial residuals" and u_s "total residuals." The system expressed in the above form is referred to as the "forecast reduced form." Let a^* be a vector containing all nonzero, nonunit elements of estimated matrix A. Then, since $F = -B^{-1}C$, each element f_{ij} of F is a function of the elements of a^*. A forecast of the ith endogenous variable (Y_{iF}) can be calculated from the ith equation of the above "forecast reduced form" system. By analogy with the single equation development, the estimated forecast variance is the sum of two components and may be written, $S^2(Y_{iF}) = Z_F[S(f_i)]Z_F' + S_{si}^2$, where Z_F is the vector of "known" values of the predetermined variables in the forecast period, $S(f_i)$ is the estimated covariance matrix of the estimated coefficients f_i, and S_{si}^2 is the estimated variance of "total residuals" μ_{si} in the ith equation. In the procedure outlined, S_{si}^2 is obtained directly from the expression, $S_{si}^2 = (1, -f_i)[M_{x_ix_j}](1, -f_i)'$, where $(1, -f_i)$ is the coefficient vector and $M_{x_ix_j}$ is the moment matrix of variables in equation i. The elements of $S(f_i)$ are obtained from, $S(f_i) = [\partial f_i/\partial a^*]S(a^*)[\partial f_i/\partial a^*]'$, where $\partial f_i/\partial a^*$ is the Jacobian of the coefficients f_i with respect to a^*, and $S(a^*)$ is the covariance matrix of the estimated structural coefficients a^*. The elements of $S(a^*)$ are estimated from the negative inverse of the matrix of second order partial derivatives of the logarithmic likelihood function. The elements of both $[\partial f_i/\partial a^*]$ and $S(a^*)$ are evaluated at the point of sample estimates a^*.

In conclusion, a few remarks are offered with respect to degrees of freedom in small samples, confidence versus tolerance intervals, and reduced form equations (coefficients of the reduced form estimated directly by least squares) versus the "forecast reduced form" as a basis for forecasting. The suggested rules in the case of small samples are taken directly by analogy from the single equation case in which Markoff conditions are met and are recognized as possibly quite inappropriate to the multiequation model. A minor error is noted in the correction factor in equation 31, which, by analogy with least squares, should read $(T/T-m)^{1/2}$. IVAN M. LEE, *University of California.*

Downton, F., "Least-squares Estimates Using Ordered Observations," *Annals of Mathematical Statistics*, 25 (1954), 303–16.

Ordered least-squares estimates are obtained for a class of 2-parameters distributions of the form $f\{(x-\mu)/\sigma\}/\sigma$. The general expressions for the quantities to evaluate the estimates of μ and σ are given for this case. In any special case, substitutions can be made in the general formulas to give the particular estimates. The ordered least-squares estimates of 2 distributions, viz. the rectangular and the right triangular, are given as special cases. The expected values, the variance matrix, the estimates and their variances are calculated for the latter distribution for samples up to size 10. Furthermore, the single parameter system of the form $f(x/\lambda)/\lambda$ is considered and from this is derived the ordered least squares estimate of λ, and the expressions for the quantities needed to calculate the estimate. In the case of the exponential distribution $f(x) = e^{-x/\lambda}/\lambda$, the estimate of λ is the sample mean. A Pearson Type III distribution, depending upon a single dispersion parameter, is discussed and the ordered least-squares estimate turns out to be identical with the maximum likelihood estimate. A. E. SARHAN, *University of North Carolina.*

Durbin, J., "Some Results in Sampling Theory when the Units Are Selected with Unequal Probabilities," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 262–69.

A rule is given for calculating estimates of sampling error that can be applied to multistage designs of any degree of complexity

and which is an extension of the rule given by Yates for the case of equal probabilities of selection. The relation between the theories of sampling with and without replacement is discussed and two approximate procedures are described which are easy to apply but lead to slight overestimates of the sampling error. T. S. RUSSELL, *Virginia Polytechnic Institute.*

Epstein, B., and Sobel, M. "Some Methods Relevant to Life Testing from an Exponential Distributiou," *Annals of Mathematical Statistics*, 25 (1954), 373–81.

The authors considered that they have N items for life testing divided into k set S_j, each containing n_j, based on the two parameter exponential distribution $(1/\theta)e^{-(x-A)/\theta}$, $A \leqq x \leqq \infty$. Each set is observed until the first r_j failures occur. Three different cases are considered according as the n_j items have a common known or unknown A_j, or the N items have a common unknown A. Some preliminary lemmas and corollaries are given concerning the r-ordered observations out of n based on the given exponential distribution. The maximum likelihood estimate of θ in the three cases is obtained. The article shows that the random variable $2R\,\hat{\theta}/\theta$ (R observed x's) is distributed as $\chi^2\,(2R)$, $\chi^2(2R-2k)$ (k sets), and $\chi^2(2R-2)$ in the three cases respectively. For the case in which the N items have a common unknown A, the confidence limits for θ not involving A, and for A not involving θ, have been worked out. They show that the three cases are equivalent to assuming that the life distribution in the various sets will plot in the form of straight line (s) whose slope (s) can be estimated by their results. A. E. SARHAN, *University of North Carolina.*

Gartaganis, Arthur J., "Autoregression in the United States Economy, 1870–1929," *Econometrica*, 22 (1954), 228–43.

Correlograms for 1 to 6 lags were obtained for each of 83 economic series taken from Burns (*Production Trends in the United States Since 1870*). The series were classified in agriculture, mining, and manufacturing sectors, also similar to those defined by Burns. The time series covering the period 1870–1929 were broken into two periods (1870–1913 and 1914–1929) for analysis and comparison of autoregressive structure. Mean correlograms were constructed for each sector and for each period. Results of approximate tests, based on "Student's" t, for differences in mean autocorrelations between time periods within sectors and between sectors within time periods for each lag are reported. Significant differences, "with minor exceptions," are reported between time periods within segments. A few significant differences appear between segments within time periods at several lags, but here the differences are not so marked. Autocorrelation coefficients for a few series with and without trend adjustment (deflation by population series) are tabled for comparison.

Although autoregressive structures are not calculated, it is asserted that graphic appraisal of the series during the 1870–1913 period indicates that they are evolutive. For the 1914–1929 period, autoregressive structures are determined for the agriculture and mining sectors from the mean autocorrelations calculated. The roots of the characteristic equations of these structures suggest that the autoregressive systems are stationary. Because of "heterogeneity" the manufacturing sector is disregarded in this latter analysis. From the mean autocorrelations and the autoregressive coefficients, approximations to mean autocorrelations for additional lags are obtained for agriculture and mining. The correlograms dampen and finally vanish. This suggests either a moving average or autoregressive type of structure. The author ventures the opinion that the structures are autoregressive. The author's main conclusions are summarized as follows: "(1) The autoregressive structure of the economy for the period 1870–1913 differs from that of the period 1914–1929 period. (2) Orcutt's hypothesis (*Journal of the Royal Statistical Society*, Series B, 10: 1, 1948) that Tinbergen's series of the 1919–1931 period can be considered as a sample drawn from a population having the autoregressive structure, $y_t=1.3y_{t-1}-.3y_{t-2}+\epsilon_t$, is not empirically substantiated by our sample." The alternative hypothesis is stated that for the similar period, 1914–1929, the American economy can be considered as a population having different underlying autoregressive structures. Two of these have been estimated in this paper. They are for agriculture and mining, respectively: $x_t=.2974x_{t-1}-.0240x_{t-2}-.0561x_{t-3}-.0192x_{t-4}+.0463x_{t-5}-.0936x_{t-6}+\epsilon_t$, and $x_t=.3654x_{t-1}+.2582x_{t-2}+.1384x_{t-3}-.2553x_{t-4}-.0798x_{t-5}+.1831x_{t-6}+\epsilon_t$. IVAN M. LEE, *University of California.*

Gulliksen, H., "A least-squares solution for successive intervals assuming unequal standard deviations," *Psychometrika*, 19 (1954), 117–39.

A least squares solution for the scale values obtained by using the method of

successive intervals for the basic observational data is derived. The theoretical solution depends upon solving simultaneously for the scale values (m_i), the discriminal dispersions (s_i), and the category boundaries (t_g) which will minimize the quantity

$$(1/b^2)\Sigma_i\Sigma_g(s_i z_{ig}+m_i-t_g)^2,$$

where z_{ig} is a normal deviate corresponding to an observed proportion and b is an arbitrarily assigned standard deviation for t_g.

Numerically the direct least squares solution is laborious; methods for simplifying the computations are presented. A series of numerical examples compare the relative accuracy of scales obtained from various computational procedures. B. J. WINER, *University of North Carolina.*

Gurland, John, "An Example of Autocorrelated Disturbances in Linear Regression," *Econometrica*, 22 (1954), 218–27.

The author investigates the loss of efficiency of estimators of the regression parameters when there are certain types of specification bias concerning the disturbances. Let $y_t=\xi_t+u_t$ $(t=1, 2, \cdots, n)$, where ξ_t, the expected value of y_t, is a linear combination, $\xi_t=\theta_1x_{1t}+\theta_2x_{2t}+\cdots+\theta_{k-1}x_{k-1,t}+\theta_k$ of the unknown parameters θ, and $Eu_t=0$. The covariance matrix of disturbances u_t (Ω) consists of elements $Eu_tu_j=\sigma^2\omega_{tj}$. It is assumed that the x's are "fixed" variates and that the elements ω_{tj} are known. The disturbances are assumed to be generated by a first-order Markoff process, $u_t-\rho u_{t-1}=v_t$, where $Ev_tu_{t-1}=0$, $(t=-N+1, \cdots, 1, 0, 1, \cdots, n)$, $Ev_t=0$, $Ev_tv_{t'}=0$ $(t\neq t')$, $Ev_t^2=\sigma^2$, $(t=-N, -N+1, \cdots, -1, 0, 1, 2, \cdots, n)$. The "initial value" u_{-N} is defined by $u_{-N}=\delta v_{-N}$, where the value of δ is selected arbitrarily. Define $g_N^2=1+\rho^2+\rho^4+\cdots+\rho^{2N+2}\delta^2$. If the true values of ρ and δ are known, the best linear unbiased estimates of the parameters θ may be obtained by means of the transformation $s_t=y_t-\rho y_{t-1}$, $\eta_t=\xi_t-\rho\xi_{t-1}$, $z_{ht}=x_{ht}-\rho x_{h,t-1}$ $(t=2, 3, \cdots, n;\ h=1, 2, \cdots, k)$, with $s_1=y_1/g_N$, $\eta_1=\xi_1/g_N$, and $z_{h1}=x_{h1}/g_N$ and solving the k linear equations $\partial/\partial\theta_h[1/g_N^2\ (y_1-\xi_1)^2+\Sigma_2^n\ (s_t-\eta_t)^2]=0$. If incorrect values are used for ρ or δ, or both, the estimates, although unbiased, will no longer be "best" in general.

Cochrane and Orcutt (*Journal of the American Statistical Association*, 44: 245, 1949) recommend neglecting the first term in the above set of k linear equations, claiming this is justified if the true value of ρ is close to 1. Assuming that the true value of ρ is known, the author investigates the loss of efficiency from unjustifiably omitting the term $1/g_N^2\ (y_1-\xi_1)^2$, that is, assuming g_N extremely large when in fact it is not. For the case $k=2$, the author derives the limiting expression for the joint efficiency of the estimated regression parameters under the incorrectly specified g_N (denoted by g^*). It is then shown that there exist values of x_1 and g^* in one case and values of ρ and g^* in another for which the efficiency is arbitrarily close to zero. Limiting expressions derived hold also in the case of evolutionary series for u_t. Three interpretations of the assumption that g^* is very large are given.

Also investigated is the possible loss of efficiency in assuming u_t to be a stationary process when, in fact, it has an initial fixed value $u_{-N}=0$. From the expression derived for joint efficiency, it is concluded that this incorrect specification could be a source of the considerable loss of efficiency of the estimates obtained by Cochrane and Orcutt from their series designated by (B).

In an appendix, the joint efficiency of estimated regression parameters is derived for the general case of incorrectly specified disturbance covariance matrix. A minor notational omission appears on page 226 where, in the covariance and joint efficiency expressions, Ω should read Ω^0. IVAN M. LEE, *University of California.*

Guttman, L., "Some Necessary Conditions for Common-Factor Analysis," *Psychometrika*, 19 (1954), 149–61.

One of the fundamental problems in common-factor analysis is: given a matrix of sample intercorrelations R having units in the diagonals, to find a diagonal matrix U_i^2 such that $G_i=R-U_i^2$ is a Grammian matrix of minimum rank. Three theorems which give lower bounds on the rank of G are developed.

Let s_i be the number of latent roots of G_i which are greater than or equal to unity. For any matrix U_1^2 leaving G_1 Grammian, the minimum rank of G_1 is shown to be greater than or equal to s_1. If U_2^2 has as its jth element the multiple correlation of variable j with all other variables, then the resulting G_2 will have minimum rank equal to or greater than s_2. If U_3^2 has as its jth element the highest zero order correlation with the other variables, then the resulting G_3 will have minimum rank equal to or greater than s_3. The three lower bounds for G can be ordered as follows: $r\geqq s_2\geqq s_3\geqq s_1$. In practice s_3 will generally be the

simplest lower bound to compute. B. J. WINER, *University of North Carolina.*

Hotelling, Harold, "New Light on the Correlation Coefficient and its Transforms," *Journal of the Royal Statistical Society,* Series B, 15 (1953), 193–232.

This paper presents parts of the derivation of the distribution of r in a new form. A method of obtaining the probabilities associated with the elementary case of the population correlation ρ being zero, for large or for small samples is given without the use of tables other than of logarithms. For $\rho \neq 0$ the slowly convergent series of incomplete beta functions for the probability integral given by Pearson is replaced by a rapidly convergent series of such functions.

The moments of r about ρ and the moments of z are calculated by a new method. This paper also examines the possibility of improvement over the use of z. Certain points in the mathematical theory of correlation coefficients are simplified to make more feasible their inclusion in future courses and textbooks. This author uses n for degrees of freedom, thus several of the derivations and formulas associated with the correlation distributions appear slightly simpler than in terms of sample number. CLYDE Y. KRAMER, *Virginia Polytechnic Institute.*

Bryon, A. Hughes, "Methods for Analyzing and Interpreting Physical Measurements of Groups of Children," *American Journal of Public Health,* 44 (1954), 766–74.

The article points out how three statistical techniques can be valuable tools in analytical surveys of physical measurements of groups of school children. First, analysis of covariance should be used to adjust for variables (e.g. age, height) which can be measured but not often controlled in sample selection. The second tool discussed involves consideration of statistical methods developed in connection with bioassay problems. The author indicates how these methods are applicable in describing the degree of sexual maturation of a group of children. The third technique is a multivariate mathematical model to be used in studying the joint effect of several growth factors. In the appendix, the author mentions some nonsampling errors that frequently are not taken into consideration by workers in anthropometric and nutrition research. BERNARD G. GREENBERG, *University of North Carolina.*

King, E. P., "Probability Limits for the Average Chart When Process Standards Are Unspecified," *Industrial Quality Control,* 10 (1954), 62–64.

New control limits for $\bar{\bar{X}}$ are presented such that

$$\text{Prob}\left\{\frac{\max|\bar{X}_i-\bar{\bar{X}}|}{\sigma/\sqrt{n}}\leq k_m\right\}\cong .95$$

where m is the number of subgroups, n is the subgroup size and $i=1, 2, \cdots, m$.

The control limits proposed are $\bar{\bar{X}}\pm C\bar{R}$ where

$$C=k_m/(\sqrt{n}d_2).$$

These limits provide "short run" $\bar{X}$ charts with approximately a constant type I error. A graph of C factors is included for subgroup sizes or 2, 3, 4, 5, and 10 and for sample numbers of 3 through 25. The C factors presented are approximate. GERALD J. LIEBERMAN, *Stanford University.*

Klein, L. R., and Mooney, H. W., "Negro-White Savings Differentials and the Consumption Function Problem," *Econometrica,* 21 (1953), 425–56.

Data from the 1947, 1948, 1949, and 1950 Surveys of Consumer Finances form the basis for the analyses summarized in this paper. Survey schedules classified by region, race, and income class suggest: (1) for the North, the ratio of savings to disposable income for comparable income classes is higher for Negroes than for whites over the entire range of income and (2) for the South the savings-income ratio is higher for Negroes than whites in the lower income classes while the relative positions are reversed in the higher income classes. The Survey data for the North are in line with relations suggested by the Consumer Purchases Study of 1935–36. The analyses summarized here bear on propositions advanced by the present authors and others previously concerning the factors "explaining" the racial differential in savings behavior and the implications of racial differentials in the construction of aggregative consumption functions.

In an early section mean residuals, by racial groups, from regression equations "explaining" the savings-income ratio are presented. Separate equations for homeowners and nonhomeowners were calculated from data supplied by an urban sample of the 1949 Survey. Data from a larger sample of nonfarm, nonbusiness spending units served as the basis for another equation in which the characteristic homeownership, was represented as a sepa-

rate variable in the analysis. Measures used directly or in construction explanatory variables were: disposable income, liquid asset holdings, number of persons in spending unit, age of spending unit head, and lagged disposable income. The residuals from regression in each equation were averaged for Negroes, whites and others. The mean of residuals for Negroes in each equation was positive and larger than for whites. The results were presented as suggestive, recognizing that the racial differences in mean residuals are not statistically significant.

Another main section of the paper reports results of variance analyses of mean savings-ratio deviations. For each regional-racial group, the deviation of each spending unit's savings-income ratio from the mean ratio of its income class was calculated. In selected analyses, year served also as a separate variable, while in others data for the four years were combined. Additional variables are then introduced on the basis of which the savings-ratio deviations are further classified. The mean of the deviations falling in each cell of the resulting multiple cross classifications is the random variable analyzed in a factorial design with one observation per cell.

Among the additional variables on which classifications for the several variance analyses are based are: (1) liquid asset holdings, (2) past income change, and (3) job security. Significant main effects of the first two of the above variables are reported as well as significant interactions of one or more of these variables with race, region, and/or disposable income. For several of the tests, data giving rise to significant interactions are reproduced to facilitate interpretation. Finally, reference is made to variance analysis results with certain other variables, although the results are not reported in detail.

A supplemental device employed in the paper is the presentation of percentage distributions of spending units with respect to the several variables introduced by regional-racial-income classes. A brief section presenting and discussing the implications of such a percentage distribution with respect to credit *use* appears in a final section of the text of the paper. IVAN M. LEE, *University of California.*

Irwin, J. O., "On the Transition Probabilities Corresponding to Any Accident Distribution," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 87–89.

From any known distribution of accidents in a fixed exposure time T, the expected number of other accidents sustained by a person who has had x accidents is shown to be the ratio of the $(x+1)$th to the xth factorial moment of the distribution. The limiting value of this ratio when the exposure time tends to zero gives the transition probabilities. From the form of the frequency distribution the transition probabilities are derived. G. L. EDGETT, *Virginia Polytechnic Institute.*

Lukacs, Eugene, "On Strongly Continuous Stochastic Processes," *Sankhyā*, 13, Part 3 (1954), 219–28.

The first theorem is concerned with the normality of increments of a strongly continuous stochastic process. The proof makes use of the ϵ, δ definition of strong continuity. Various properties of strongly continuous processes are then derived and used in the proofs of theorems 2 and 3. Theorem 2 states necessary and sufficient conditions for a stochastic process to be a Wiener process and the sufficiency of the conditions is demonstrated. Theorem 3, the last theorem of the paper, shows that the variance of a strongly continuous process with independent increments need not be independent of the time t. F. S. MCFEELY, *Virginia Polytechnic Institute.*

Masuyama, Motosaburo, "Mathematical Note on Area Sampling," *Sankhyā*, 13, Part 3 (1954), 241–42.

The author gathers together some results of integral geometry due to Poincare, Crofton, Blashke, Santalo and his own work in order to draw attention to the possibilities of application to statistical problems of area sampling. R. J. TAYLOR, *Virginia Polytechnic Institute.*

Matthai, Abraham, "On Selecting Random Numbers for Large Scale Sampling," *Sankhyā*, 13 (1954), 257–60.

Random numbers in small scale work may be selected without much regard to cost considerations; but large scale work requires a method which reduces the labor of selection.

In an example cited it is necessary to choose a random sample of 800 out of 70,000. The selection rule is to assign ten digits in a random number table to each five digits to be selected, and to take the last four digits prefixed by the first digit to the left less than 7; To illustrate:

67	6345	0912 gives 50912
25	2987	4391 gives 24391

In a second case, one must select a set of random numbers less than 2853. A similar

method of selection has an expected rejection rate of 4.9%, as compared with 71.5% in the method of rejecting all four digit numbers greater than 2853 and with 14.4% in the method of dividing by 3000 and taking remainders.

The appropriateness of the method is established by χ^2 tests. A. N. POZNER, *Virginia Polytechnic Institute.*

McGill, W. J., "Multivariate Information Transmission," *Psychometrika*, 19 (1954), 97–116.

A model for handling multidimensional contingency tables in terms of information theory is developed. The method of analysis used is analogous in some respects to the analysis of chi-square into its components. The sampling distributions of some of the statistics that are computed in the course of the analysis, particularly those concerned with interaction effects, have not been tabulated. When such tables become available, the method developed should provide the research worker with a useful analytic tool.

The information transmitted from two inputs, u and v, to an output, y, is definedto be

$$T(u, v; y) = T(u; y) + T(v; y) + A(uvy),$$

where $T(u; y)$ and $T(v; y)$ represent the bivariate transmission in bit units and $A(uvy)$ represents the interaction effect. One measure of the interaction effect is shown to be

$$A(uvy) = T_v(u; y) - T(u; y),$$

where $T_v(u; y)$ is the average information transmitted between u and y for constant value v. Extension of this model to the general multivariate case involving several orders of interaction is direct.

A numerical example analyzing output information into the equivalent of main effects and interaction effects is worked out in detail. Significance tests of the interactions involve the sampling distribution of differences between chi-square variables. Although the density function of this distribution has been derived, tabled values of this function are said to be lacking. Approximate distributions for other statistics involved in multivariate transmission have been developed. Certain of these distributions useful in testing main effects are given. B. J. WINER, *University of North Carolina.*

Moran, P. A. P., "The Estimation of the Parameters of a Birth and Death Process," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 241–46.

The estimation of $\lambda(\lambda+\mu)^{-1}$ is considered for a simple birth and death process. It is equivalent to the problem of estimating the probabilities of steps to the right and left from an observed realization of a random walk which has one absorbing boundary and which is terminated, if necessary, after a preassigned number of steps. The properties of various estimators are considered. T. S. RUSSELL, *Virginia Polytechnic Institute.*

Olkin, I., and Roy, S. N., "On Multivariate Distribution Theory," *Annals of Mathematical Statistics*, 25 (1954), 329–39.

The authors developed a matrix method of handling a large class of multivariate distribution problems including in particular those for which the Wishart distribution is not available (e.g., the case of a sample of N observations from a p-variate normal population with $p > N-1$). Two techniques are used for evaluating the Jacobians of certain transformations. The first is applied to obtain the joint distribution of the rectangular coordinates. The second is applied to obtain the joint distribution of the roots of a determinantal equation. A. E. SARHAN, *University of North Carolina.*

Podder, K. C., "On the Punched Card Method in Smoothing for Age Bias in Census Returns," *Sankhyā*, 13, Part 3 (1954), 261–66.

A method of smoothing for age bias in the preparation of the 1951 census Tables of India using Hollerith computing equipment is described. A table showing the method used and a specimen working table are given. The actual machine operations are described by use of a cycle chart and Control Panel wiring diagrams. R. J. TAYLOR, *Virginia Polytechnic Institute.*

Psychological Research Wing, "Multiple Factor Analysis of Personality Ratings in Services Selection Boards," *Sankhyā*, 13 (1953), 17–26.

The purpose of the investigation reported in this paper was the study of the functional unities underlying the checking of qualities on a rating scale used by Indian Army Selection Boards for the selection of officer candidates, and as such is a good example of a complete factor analysis. A sample of 418 boys, each studied in relation to 21 qualities, was taken, and the resulting data was subjected to a complete centroid factor analysis, utilizing Tucker's criterion for the stopping rule. Three factors were extracted which were felt could sufficiently explain the inter-correlations. The resulting matrix was rotated by the method of extended vectors, and the three primary

factors were obtained and identified as: (1) Intellectual Factor, (2) Social Factor, and (3) Dynamic Factor. H. C. SWEENY, *Virginia Polytechnic Institute.*

Sengupta, J. M., "Some Experiments with Different Types of Area Sampling for Winter Paddy in Giridih, Bihar: 1945," *Sankhyā*, 13, Part 3 (1954), 235–41.

The object was to study the relative efficiencies of different sampling units, with variations in the method of enumeration, for the estimation of acreage under winter paddy. Three different methods, using two different types of sampling units, were used, being discussed and compared with respect to bias, cost and efficiency. Their advantages and disadvantages are given. DANIEL ZAKICH, *Virginia Polytechnic Institute.*

Sharma, O. C., "Factor Analysis of Technical Trades and Educational Examination Marks of the Aircraftsmen of the Indian Air Force," *Sankhyā*, 13 (1953), 27–34.

A factor analysis was made on the results of seven final examinations taken by 75 aircraftsmen training for the Radio Telephone Operators and Telegraphists trade in the Indian Air Force. Five of these examinations were 'trade' tests, the other two being educational tests (mathematics and science). The factor analysis was done using two different techniques: (a) the Centroid Method, and (b) the Method of Principal Components. In each case, the analysis was carried out to three factors. These three factors accounted for 55.7% of the total variation in the Centroid Method and 73.5% of the total variation in the Method of Principal Components. A stopping rule by Burt was used in each case. The resulting factor matrices were rotated by means of the Method of Extended Vectors to verify the existence of simple structure. Both methods demonstrated the same factor pattern. Three group factors were obtained and identified as: (1) Clerical Ability Factor, (2) Number Ability Factor, and (3) Technical Skill Factor. H. C. SWEENY, *Virginia Polytechnic Institute.*

Singer, K., "Application of the Theory of Stochastic Processes to the Study of Irreproducible Chemical Reactions and Nucleation Processes," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 92–106.

Let $n_1, n_2, \cdots, n_r$ symbolically denoted by the vector n be the number of the different molecular species $1, 2, \cdots, r$, in a reacting system of constant volume. Suppose the system is subject to random variation and the composition is characterized by $P(n; t)$, the probability that the system has the composition n at the time t. Difference-differential equations involving probabilities of changing from one composition to another in a given time are derived and studied as are also equations involving "first passage" times and "recurrence" times. Several applications to the study of chemical reactions are given. PAUL N. SOMERVILLE, *Virginia Polytechnic Institute.*

Singh, R. P., and Nagar, D. N., "A Study on the Growth of Population in Rajosthan," *Sankhyā*, 13 (1953), 39–42.

Some data for the Rajputana states is taken from the census reports of the period 1901–1941 and studied with regard to the number of married females in different age-groups, reproduction according to age groups, distribution of married females of reproductive age, average number of children born, the number survived, sex ratio and increase in population. R. L. WINE, *Virginia Polytechnic Institute.*

"The National Sample Survey: General Report No. 1," *Sankhyā* 13, Parts 1 and 2 (1953), 47–218.

This paper reports on the National Sample Survey of India covering the period October 1950 to March 1951. The survey was conducted to supply reliable statistics relating to production, consumption and other aspects of economic and social life in India. Data were obtained on size of rural households, per capita consumer expenditure in rural areas, expenditure on food, expenditure on clothing and head and footwear, and medical and ceremonial expenses. Appendix 2, pp. 136–198, contains tables reporting on the data collected under the above noted general headings. Appendix 3, pp. 197–214 contains facsimile field schedules.

The design of the survey is discussed and some notes are included relating to changes to be made in the second round of sampling. Different methods of selecting the sampling units were adopted in different parts of the country and the probability of being included in the sample differed from region to region. Sampling units were selected in two stages: first the villages were selected after suitable stratification; within each sample village all or a subsample of 30 households, whichever was less, were stratified into agricultural and nonagricultural classes and sample households were

then selected at random from each of these strata. RALPH A. BRADLEY, *Virginia Polytechnic Institute.*

Whittle, P., "The Analysis of Multiple Stationary Time Series," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 125–39.

The author extends his earlier methods on the application of the least square principle to the analysis of a single stationary time series to its application in the analysis of a multiple series. For a purely nondeterministic stationary multiple process the least square estimation equations are derived. For a normal process the asymptotic covariances of the parameter estimates are calculated. The methods developed are illustrated by the testing of a sunspot model. G. L. EDGETT, *Virginia Polytechnic Institute.*

Wold, Herman O. A., "Causality and Econometrics," *Econometrica*, 22 (1954), 162–77.

Following a few general remarks on the concept of causality, definitions are proposed which are considered useful in problems involving relations between variables. Attention is given both to the nonstatistical (exact relations) and statistical points of view, but the discussion centers primarily on the latter. Consider the general relation $y=f(x_1, \cdots, x_h)+z$, where z represents the disturbance term. In a controlled experiment the x's represent control variables and y the effect variable. With improper design and analysis this may be interpreted as a causal relation. In the case of nonexperimental observations (for example, econometric analysis of time series data) a relation like that above is defined as causal if "it is theoretically permissible to regard the variables as involved in a fictive controlled experiment with $x_1, \cdots, x_h$ for cause variables and y for effect variable." Proper specification and analysis then permits the causal interpretation. Within the framework of the definition proposed a few simple, illustrative economic models are discussed from the econometric point of view. The causal interpretation in the illustrative relations is discussed primarily within the framework of recursive systems. Employing the concept of a *link set* the author extends the recursive model to include the case where one or more effect variables in a given "link" (endogenous variables at point t) are jointly causally explained by variables in "previous links" (lagged endogenous and lagged or current exogenous variables). IVAN M. LEE, *University of California.*

Woolsey, Theodore W., "On the Use of Sampling in the Field of Public Health," *American Journal of Public Health*, 44 (1954), 719–40.

The American Public Health Association Statistics Section had its Committee on Sampling Techniques prepare this valuable article on the uses of sampling for public health workers. The discussion is broad enough to include applications of sampling in all fields. The manuscript describes when and how sampling may be put to advantage, its reliability, and also those situations for which sampling is not a help. Probability sampling is discussed and several illuminating illustrations are presented. In the appendix, a selected bibliography on sampling is given with a list of recent references in which probability sampling was used to solve public health problems. BERNARD G. GREENBERG, *University of North Carolina.*

Yates, F., and Grundy, P. M., "Selection Without Replacement from Within Strata with Probability Proportional to Size," *Journal of the Royal Statistical Society*, Series B, 15 (1953), 253–69.

In sampling without replacement with probability proportional to size, the usual formula for estimation of a stratum variate by weighting the units in inversion proportion to the size of the units is biased. Numerical examples are given to show that the bias is small. A formula for unbiased estimates is given, which, however, for samples of size greater than two involves considerable labor.

The bias in the ordinary formula for the estimation of error is investigated and also is found to be small. An unbiased estimate of error is given which is shown to be more efficient that that given by Horvitz and Thompson.

A method of revising size measures so that, with the usual method of selection, the true total probabilities of selection are proportional to the original size measures is given for samples of size 2.

Criticism is given of the practice of selection of successive members of a sample with sets of probabilities chosen solely so that the total probabilities shall be proportional to the original size measures. PAUL N. SOMERVILLE, *Virginia Polytechnic Institute.*

BOOK REVIEWS

Sexual Behavior in the Human Female. *Alfred C. Kinsey, Wardell B. Pomeroy, Clyde E. Martin, and Paul H. Gebhard.* Philadelphia and London: W. B. Saunders Company, 1953. Pp. 842. $8.00.

See review article by Dorothy S. Brady, on pages 696–705.

Statistical Method in Industrial Production. Thirteen Papers plus Foreword by A. Bradford Hill given at a Conference held by the Industrial Applications Section of the Royal Statistical Society in Sheffield in 1950. London: 1951. Pp. iv, 89. 7s 6d.

LLOYD A. KNOWLER, *State University of Iowa*

"EARLY Experiences of Statistical Quality Control in a Pottery," by Arthur G. Ellis, is a case history of two applications in the manufacture of pottery—pint weight of slip and dry modulus of rupture. The tremendous benefits which can result in the industry from the use of statistical quality control chart techniques is indicated. Also, the importance of having missionary work at or about the foreman level is noted.

"Applications of Control Charts to Brick Manufacture," by T. G. W. Boxall, describes an application of quality control charts in an industry in which it is impossible to make big changes in the source of raw material and which is concerned with the manufacture of a cheap, mass-produced article. Through the measurement of but six bricks of about 40,000 burnt in a kiln, it was discovered that a simple modification to the feeding mechanism of the presses would result in bricks which would easily meet the specification limits set by the British Standards Institution. Another study indicated some minor changes desirable in anticipation of shrinkage characteristics. Following these two studies, the control chart techniques have been expanded so as to consider crushing strength, weight, and absorption of bricks, as well as the quality of bricks made in different machines and burnt in different kilns. Also, the technique has been expanded to compare the efficiency of various brickmaking operations and firing methods.

"Contributions of Statistics to Problems of Chocolate Manufacture," by B. Moorhouse, describes a study of the manufacture of moulded chocolate block containing a "centre" which, on completion, was showing more variation in weight than was desired. It was noted that among three main lines to bring the manufacturing process under control were efforts to reduce (*1*) variation between moulds, (*2*) variation between cake positions within the mould, and (*3*) variance between day-to-day runs. It is pointed out that the best way to acquire furrther knowledge of control chart work is to make as many applications as possible, and that conclusions drawn from experiments carry more weight when the data are examined by the statistical technique.

"Productivity Measurement in the U.S.A.," by H. Ingham, "analyses the

differences in Great Britain and the U.S.A. in attitudes to productivity, and urges that (Great Britain) should energetically engage in certain specific statistical investigations of productivity." Among the points made are the following: (1) Productivity seems to have become a national myth in the U.S.A.; in fact, it is pointed out that business men and trade unionists would become seriously concerned if the figures showed that productivity had not increased at a rate of about 3 per cent per annum. (2) The people in U.S.A. are interested in "man-hours required per unit of output." (3) In U.S.A. there is an overwhelming emphasis on the "down to earth" type of person, while the reverse seems to be true in Great Britain.

"Costing of Continuous Processes," by Philip Lyle, illustrates application of statistical methods to the determination of the average effect upon the costs of a factory, department, or process of a change in output. It is shown that a measure of the total variation amongst a series of weekly cost figures can be divided into (1) the amount of this variation which can be ascribed to changes in output, and (2) the amount due to unknown factors or "error." The knowledge of the error component enables one to predict the costs based upon various outputs. Marginal cost is discussed. It is shown that the marginal concept only applies for short-run variations, and that long-run variations take place in finite steps for which "arc cost" must be used in place of "marginal cost."

"Graphical Analysis of Variations as a Production Department Tool," by E. A. G. Knowles and C. Roseman, shows, by means of an example, how a graphical analysis of the results of a complicated experiment can be carried out by members of a production department familiar with control charts. In the example, all effects show up on the control charts, and they are made available immediately to all concerned. It is indicated that final tests of significance, such as analysis of variance, can be performed by the statistical section of an organization. A comparison of the results of the graphical analysis are made with those of analysis of variance.

"Comparative Tests in a Single Laboratory," by W. J. Youden, makes the point that "It is a fact of experience that a set of measurements made by different operators at different times or in different localities is subject to greater variation than a set of measurements made by one operator using the same apparatus on the same day. The data from an experiment with thermometers have been used to show that even as simple an operation as reading a linear scale cannot be duplicated nearly as well after a time lapse as on the same day. The paper emphasizes the necessity of including in the error of the test all sources of variation which in fact operate on the measurements and shows how a change in the design of the test may reduce the error."

"The Statistical Approach to Time Study," by D. J. Desmond, "first gives a brief historical survey of the development of Time Study and proceeds to show how the technique of rating has become a part of modern time study in many industries. The methods of selecting the normal time for a job are then discussed, and it is shown that until recently there was no ob-

jective way of determining the quality of any particular time study or measuring the accuracy with which any time study observer is working.

"A new method of analysis is then developed, based on regression analysis, which gives an objective determination of the normal time of a job in terms of the recorded times and subjective estimates of the observer. This will give an estimate of the unknown normal time, and the precision of this estimate can be calculated and compared with the results obtained by other observers studying the same job. The various defects in a study can be calculated in terms of three different parameters which establish the standard of quality of the studies of the observer. Plotting these statistics on control charts enables the observer to determine, at a glance, whether he is maintaining the quality of his work, and to see if he is achieving any improvement. A simple graphical method is described which enables him to estimate all the characteristics of his study in less time than he usually takes merely to determine his normal time.

"The method is then developed, by the analysis of variance technique, to enable any number of studies to be combined. This can lead to the establishment of a standard of quality for a group of observers, and the significance of the differences between individual observers can be examined. These differences are illustrated by the results of an experiment carried out on the floor of an assembly shop."

"Problems of Even Flow in Production," by E. D. Van Rest, deals with what are sometimes called "congestion" problems, which arise in providing a service when the need arises at random intervals of time. Such problems are frequent in industry; for example, one operator tending several machines or one machine tended by two or more operators. In fact, the problem is important in planning an even flow of work through a production process because the various accidents which are liable to occur delay progress. The particular problem considered is typified by the spinning frame of a cotton mill where one person looks after a large number of spindles. The thread of each spindle may occasionally break and need repair. The type of information required is described, as well as the use to be made of it. Similar problems have received some attention in operations research, as well as in quality control.

"Statistics Applied to Assembly Process," by G. A. Barnard, considers the tolerance limits of an assembled article as related to those of the components. In particular, consideration is given to the following types of assembly process: (*1*) random or interchangeable, (*2*) semi-random, (*3*) simple selective, and (*4*) multiple selective. The need for a mathematically trained statistician in any fair-sized plant, to form a link between the production and cost departments, is observed.

In "The Cost of Inspection," by F. J. Anscombe, "assessment of the total cost of an inspection procedure is considered, taking into account the cost of decisions made on the basis of the inspection. Simple hypothetical process curves, inspection cost curves, and decision loss curves, are described. A

numerical example of rectifying inspection is considered in some detail, and the relevance of the Dodge-Romig concepts of AOQL and lot tolerance to such problems is discussed."

"Multiple Sampling in Theory and Practice," by J. H. Enters and H. C. Hamaker, selects "from among the great variety conceivable such multiple plans . . . as can be presented to inspectors in the form of very simple instructions, in which use is made of the method of scoring proposed by Barnard for sequential sampling. For plans of this type the operating characteristics and the average sample size are computed assuming Poisson probabilities. A measure of efficiency, the inverse efficiency, is then obtained by dividing the average sample size by the sample size of a single sampling plan possessing practically an identical operating characteristic. The search for these equivalent single sampling plans is greatly facilitated by specifying the operating characteristics by their point of control, p_0, and their relative slope, h_0, defined by

$$P(p_0) = \tfrac{1}{2}$$

and

$$h_0 = -\left(\frac{p\,dP}{P\,dp}\right), \qquad p = p_0,$$

where $P(p_0)$ is the probability of accepting a lot in which the proportion of defectives is p_0. On this basis a number of multiple sampling plans are investigated. Their efficiency is compared with that of double and sequential sampling, and the influence of the crudeness of the steps and of curtailing is systematically studied. The actual number of observations is a stochastic variable the distribution of which is separately considered. In a final section the experience gained in applying multiple sampling in a factory over a period of about three years is briefly recorded."

"Sequential Analysis of Machine Performance," by B. H. P. Rivett, considers situations where the variation of a dimension of the product of a machine is sufficiently small compared with the tolerance, so that the machine setting can have a zone within which it is free to move without defectives being produced. A method is given for determining (assuming certain risks) whether the setting of a machine is in this zone. The method can be adapted to a lot-by-lot inspection scheme for acceptance of the product with reference to the mean dimension.

Research Methods in the Behavioral Sciences. *Leon Festinger and Daniel Katz, editors.* New York: The Dryden Press, 1953. Pp. xi, 660. $5.90.

DANIEL O. PRICE, *University of North Carolina*

THIS very excellent volume might more appropriately have been titled *Research Methods in Social Psychology*, for the actual title seems merely to capitalize on a new and popular term. As soon as the reader realizes that

the authors are not trying to make social psychology synonymous with the behavioral sciences, resentment dies out and the real merits of the book are more clearly seen.

Following a short introduction on The Interdependence of Social-Psychological Theory and Methods: A Brief Overview (Theodore M. Newcomb), the volume is divided into five parts: Research Settings, Procedures for Sampling, Methods of Data Collection, The Analysis of Data, and The Application of Research Findings.

Part I, Research Settings, deals with The Sample Survey, Field Studies, Experiments in Field Settings, and Laboratory Experiments. These chapters, each by a different author, are well integrated.

Part II, Procedures for Sampling, has only one chapter, Selection of the Sample by Leslie Kish. In the reviewer's opinion this is one of the best brief (65 pages) treatments of sampling that a research worker can find in the literature. It is sound and practical, even including a brief section on non-sampling errors.

The Methods of Data Collection (Part III) includes Problems of Objective Observation; The Use of Documents, Records, Census Materials, and Indices; The Collection of Data by Interviewing; and the Observation of Group Behavior. Had the book been written under the title which it now carries, we might have expected a chapter on case studies of individuals. The chapter on Problems of Objective Observation (Helen Peak) includes, among other things, comments on item analysis, comparisons of Thurstone, Likert, and Guttman scales, and discussions of validity and reliability. The chapter on The Collection of Data by Interviewing (Charles F. Cannell and Robert L. Kahn) includes not only material on the psychological basis of the interview and principles of interviewing but also material on questionnaire construction, training of interviewers, and a detailed sample interview. The chapter on Observation of Group Behavior (Roger W. Heyns and Alvin F. Zander) deals with "two principle types of observation instruments: category systems and rating scales," and deals only briefly with observational situations.

Despite the generally high quality of this volume, Part IV, The Analysis of Data, is probably the meatiest section of the book. The chapter on Analysis of Qualitative Material (Dorwin P. Cartwright) is an excellent presentation of how to develop and use a plan of content analysis or coding (the terms are used interchangeably). The Theory and Methods of Social Measurement (Clyde H. Coombs) is a chapter that gets at the very roots of social measurement, though it is so tightly written as to be quite heavy going in places. (Coombs uses the term "qualitative" in a different sense than does Cartwright in the preceding chapter.) Keith Smith's chapter on Distribution-free Statistical Methods and the Concept of Power Efficiency is, among other things, an excellent collection and presentation of distribution-free statistical methods.

The last part and chapter, The Utilization of Social Science (Rensis Likert

and Ronald Lippitt), is a good discussion of the procedures, policies, and problems involved in the application of research findings.

All chapters include good bibliographies and lots of live, illustrative material. The book will, quite properly, find a wide market as a text and reference book in research methods.

Income and Wealth: Series III. *Milton Gilbert, editor.* Papers by Milton Gilbert, Shigeto Tsuru and Kazushi Ohkawa, Richard Stone, and Kurt Hansen, Tibor Barna, S. Herbert Frankel, Frederic Benham, V. K. R. V. Rao, Daniel Creamer, Ingvar Ohlsson, and Francois Perroux, Georges Guilbaud, Jacques Mayer, Jean Albert, and Marcel Malissen. Cambridge: Bowes and Bowes, 1951. Pp. xiii, 261. Price 35s.

EARL R. ROLPH, *University of California (Berkeley)*

THIS volume contains ten papers delivered at the meeting of the International Association for Research in Income and Wealth held at Royaumont (France) in 1951. Two of the papers provide data on national income over a long period—for France since 1780 and for Japan since 1878. The detailed information these papers contain should be of especial interest to economic historians. Of the remaining papers, four apply social accounting concepts to underdeveloped areas, three deal with conceptional and theoretical topics, and one, likely to be of most interest to statisticians, is an analysis of the problem of the reliability of national income data.

Milton Gilbert maintains, persuasively in my judgment, that the reliability of a national income component can be learned only by reviewing the sources of the data and the methods of estimation employed. Meaningful numerical measures of reliability cannot be provided and attempts to do so might easily be misleading. Independent estimates do not always increase reliability because in many cases one source of data is known to be definitely superior to any other. The great differences in the quality of the data out of which national income statistics are built means in fact that national income estimates of different countries are not truly comparable, even if the conceptual differences are unimportant. This observation lends added weight to the opinion of those who were dubious of the value of an international agreement on basic income concepts, especially since it comes from one who was an important participant in those conferences.

The paper of Ingvar Ohlsson is devoted to that much discussed topic, the treatment of government activities in social accounting. There is little new information for those who have followed that literature, unless it is the resurrection of the plea for more attention to the purposes of constructing national accounts. If the plea were taken seriously, the construction of social accounts might be indefinitely postponed while debates were carried on as to what purposes are important. Presumably the purpose of intellectual work is to tell the truth as best one can regardless of how congenial or uncongenial

the results may be. Pragmatism is also the tone of a longish paper by Richard Stone and Kurt Hansen on inter-country comparisons. The authors succeed in arriving at definite conclusions, though one might wish that they would pause once in a while to inform their readers why the tests they select have relevance—why, for example, the effect on relative prices is a proper basis for distinguishing among taxes. It is hard to think of any government action that does not in fact affect relative prices.

Tibor Barna extends relativism to economic theory; apparently we must have a different set of economic theories for every country. I was surprised to learn that in France, in contrast to Great Britain, it may be proper to treat the repayment of public debt as a part of national income because in France such repayments induce increases in private expenditures whereas in Britain they do not. Mr. Barna would, I think, have some difficulty in finding a consensus among British economists that monetary policy is completely unworkable in their country. But the compiler of national income statistics need not venture into the difficult questions of monetary policy. The repayment of any debt, including a public debt, is an exchange of assets—not an income transaction—in France, Great Britain, or India. Conceptual distinctions need be kept apart from the determinants of behavior.

Of the four papers concerned with underdeveloped areas, Mr. Frankel's is mainly an elaboration of the view that it is wrong to suppose that an increase in real national income can be assumed to increase welfare. With this position one may agree or disagree, but with his insistence that the mere calculation of national income involves an implicit acceptance of certain welfare notions, I at least cannot agree. Frederick Benham in his Comments provides some sobering analysis of Frankel's rather strongly, but not always clearly, stated remarks. V. K. R. V. Rao tackles the difficult problem of international comparisons of real income. refuting the common assumption that the real incomes of less developed societies are comparatively understated because of their greater amount of household industry. One may endorse his recommendation that, with the present state of knowledge, the United Nations cease setting out figures purporting to be international income comparisons. The reader might find his remarks more convincing if he had avoided basing some conclusions on his own personal value judgments, such as that the real national income of the United States is overstated because we include the activities of the liquor business in the totals. Mr. Creamer in his paper cites chapter and verse for the advantages of having national income data for an underdeveloped area—Puerto Rico. He tempers his remarks with the warning that there are other and, in some cases, better ways to spend intellectual resources devoted to the study of underdeveloped areas than in estimating national income. His paper is informative.

Papers delivered at a conference rarely make a satisfactory book. Careful editing would have made this a smaller and perhaps a better one. It is fervently to be hoped that in any future volume of this series, an index will be provided.

Consumer Attitudes and Demand, 1950–1952. *George Katona and Eva Mueller.* University of Michigan: Institute for Social Research, Survey Research Center Publication No. 12, 1953. Pp. v, 119. Paper $1.50; cloth $2.00.

WALTER D. FISHER, *Kansas State College*

THIS empirical study reports on the buying behavior of United States families in a period of prosperity immediately following inflation. In many ways it is "an extension of research into consumer attitudes, expectations, and intentions initiated in the Surveys of Consumer Finances" (p. *iii*). It is a pioneering work, using relatively new concepts having future promise, and at the same time a workmanlike job. Although the book is thin, the material inside is meaty.

The basic hypothesis tested is that the amount of consumer spending on durable goods is influenced by certain "attitudinal" variables, including perceptions, expectations, and opinions as expressed by consumers themselves. These concepts, developed in some detail in Katona's earlier book, *Psychological Analysis of Economic Behavior*, are reviewed briefly in a theoretical chapter. Ample evidence is produced to establish this hypothesis, at least in the short run, although more attention is given to opinions of buying conditions than to actual purchases. Factors having most influence on these opinions are indicated to be consumers' perceptions of price movements in the recent past, and their evaluations of the general economic outlook in the near future.

Some of the most interesting findings concern prices and inflation. Consumers were definitely conscious of and resented the price increases of 1950 and 1951, and these attitudes affected adversely their willingness to buy. However, they did not fear inflation to the extent of making any appreciable shifts in the form of their savings from bonds to stocks, nor did they fear for the soundness of their money.

Findings are based primarily on data from four successive interview-surveys, each a sample of approximately 1000 families representing all private dwelling units in the United States, taken about six months apart with the first one in June, 1951. Each sample was independently drawn by a process of four-stage area probability sampling, using the controlled selection features developed by the Sampling Section of the Survey Research Center. An appendix table contains convincing evidence that the four samples were nearly identical in a variety of demographic characteristics such as size and occupations of the families. The interviews, approximately an hour long, contained fixed questions with no latitude given to the interviewer regarding formulation and sequence, except for occasional probes of indefinite answers.

The sample data are presented throughout in the form of percentages of responses, or of respondents, having certain attributes. Two major techniques are used: (1) answers to identical questions are tabulated separately for each time point, trends being inferred by making comparisons between findings at different times; and (2) answers to different questions—usually two at a

time—are presented in contingency tables with all time periods pooled together, and relationships inferred between two factors at a time by noting differences in certain percentages. Although no reference to statistical significance is made in the text, the reader is able to make his own judgments by use of an excellent table of sampling errors in the appendix; and, in fact, most of the findings claimed are statistically significant by conventional standards.

At times the use of coefficients of association or similar measures from the theory of attributes would have aided the reader in digesting the many arrays of percentages displayed. In some sections more use of joint relationships involving two or more independent variables would have been interesting and also more indicative of the relative importance of the various factors.

Relationship between actual purchases and the other variables could have been claimed more effectively from the time-series comparisons alone. The procedure followed of seeking to establish a relationship between reported purchases "during the last 12 months" and the opinion "this is a good (or bad) time to buy" is not convincing—for reasons which the authors recognize: first, the "last 12 months" is rather a long time in the context of this study; second, respondents would tend to rationalize recent purchases, especially since in the interviews the opinion expressed followed immediately the statement of purchases.

The authors advance also a second more ambitious hypothesis: that the use of attitudinal variables significantly improves knowledge and predicting ability over what can be done by using non-attitudinal variables alone. "It is claimed here that the use of functional relationships between consumer attitudes (as well as traditional financial variables) and spending will increase the probability of correct predictions" (p. 58). The present volume alone does not establish this claim, and does not seem designed to do so. No comparisons are made between attitudinal and non-attitudinal variables as predictors, nor between non-attitudinal predictors as used alone and as used along with attitudinal ones. Moreover, no empirical evidence is presented that would contradict a hypothesis that all attitudinal variables are ultimately dependent on or caused by non-attitudinal ones. The possibility of admitting such a view seems to be entertained by the authors when they state: "Changes in attitudes are rarely fortuitous. They are dependent on developments which induce people to restructure their thinking" (p. 57).

It may well be found useful, in the formulation and testing of models of economic behavior, to introduce variables that cannot be classified clearly as "attitudinal" or "non-attitudinal." Indeed, one of the most significant variables found in this study—the frequency of the opinion that prices went up in the recent past—conceptually "does not represent an attitude toward economic matters in the strictest sense of the term, registering instead a perception which might influence attitudes" (p. 46). Further research will determine whether other such borderline cases exist, and will also clarify the

nature of the causal interaction between the variables—psychological and otherwise—that enter into economic fluctuations and development.

The real contribution of this book lies in its emphasis on matters that have been somewhat neglected in economic analysis—especially the importance of consumer demand in business cycles, and the role of psychological variables; and also in its demonstration of the feasibility of representing such variables by answers to questions in interview-surveys. Moreover, it helps fill a great current need for more empirical work and more interdisciplinary research in the social sciences.

Cardano, The Gambling Scholar. *Oystein Ore.* Princeton, New Jersey: Princeton University Press, 1953. Pp. xiv, 249. $4.00.

MEYER DWASS, *Northwestern University*

THIS is a story of scholarship in the fascinating and fantastic Renaissance. The scholar, Cardano, is presented in a light of sympathy and understanding, which for him, is an aura distinctly new. There is, for instance, the matter of Tartaglia and the cubic. E. T. Bell gives us a typical report (Development of Mathematics, 2nd edition, McGraw-Hill, p. 117): "Cardan . . . whose name ornaments the solution of the cubic in every intermediate textbook on algebra, obtained the solution from Tartaglia under promise of secrecy and published it as his own in the *Ars Magna* (1545)." We get from Ore a distinctly new slant on what has become an old party line: Sometime before 1515, an Italian professor, Scipione del Ferro, invented a method to solve the equation $x^3+ax=b$. As was then the custom, the result was buried in secrecy. A favorite pastime of Renaissance academicians was a type of quiz contest with a heavy jackpot as well as points toward academic advancement for the winner. Hence, results such as Ferro's were not as a rule published, but were kept as secret weapons for these public disputes. It was in just such a public dispute, years later, in 1535, that Tartaglia rediscovered the method. Cardano, a physician of universal interests, was writing what he hoped would be the complete algebra of his day. Cardano succeeded in wrangling the result from Tartaglia, but only under the frustrating oath that it never be disclosed or published. This was in 1535. In the ten years that followed Tartaglia's rediscovery, still others rediscovered the method. Moreover, Cardano and a pupil succeeded in finding methods for dealing with more general forms of the cubic. What is more important, Cardano unearthed Ferro's original result and priority. Thus, Cardano felt himself relieved of oath and duty. His *Ars Magna*, published in 1545, contained the method, a statement that it was given to him by Tartaglia, and also a statement allocating priority to del Ferro. Cardano stands vindicated.

What should be of greater interest to statisticians is Cardano's virtually overlooked role in the early history of probability. Ore promotes the thesis that the father was not Pascal but Cardano. Cardano was a passionate gambler and it was inevitable that his mathematical interests should lead

him to theoretical speculations on the laws of chance. He had the miserable habit, however, of writing down speculations on little scraps of paper, jotting down improvements, revisions, and random thoughts as they came to him. Eventually the scraps were published with insufficient rewriting or editing—a collection of facts and ridiculousness. Ore dissects these hitherto undissected writings in a triple role of mathematician, classicist, and detective. Among his conclusions are the following: Cardano understood and formulated the definition of probability of an event in terms of equally likely cases. He used this to compute correctly many of the probabilities for dice and other games. He also succeeded in computing many probabilities incorrectly. His main device in the latter would in modern terms read something like, $P(A \text{ or } B \text{ or } C \text{ or } \cdots) = P(A)+P(B)+P(C)+ \cdots$. However, he fully realized that this was an approximation which was often quite unsatisfactory. He also evolved the "power law," that the probability of n occurrences of an event A in n independent trials is $P^n(A)$.

In this review (as in the book) are emphasized two highlights of Cardano's life—the cubic and probability. But Cardano lived, loved, invented, gambled. suffered, and died. Ore describes all this in a crisp and readable style. This is a book I recommend.

Gamma Globulin in the Prophylaxis of Poliomyelitis: An evaluation of the efficacy of gamma globulin in the prophylaxis of paralytic poliomyelitis as used in the United States 1953. *Public Health Monograph no. 20.* Report of the National Advisory Committee for the Evaluation of Gamma Globulin in the Prophylaxis of Poliomyelitis, Public Health Service Publication No. 358, U. S. Department of Health, Education, and Welfare. United States Government Printing Office, Washington: 1954. For sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.—$1.25, pp. vi+178.

THE 1953 study reported here is not to be confused with the 1954 vaccine trials for the control of poliomyelitis.

Experiments by Hammon and associates based on 12 gamma-globulin-inoculated cases and 16 gelatin-inoculated cases suggested that gamma globulin might be useful in modifying the severity of poliomyelitis, or even in preventing it (p. 3).

A national study during 1953 was conducted by the Communicable Disease Center of the Public Health Service, planned and guided by a National Advisory Committee (p. 1); during the 1953 summer 235,000 children were inoculated in cities and communities where there were outbreaks of poliomyelitis. It is said (p. 1) that "the records of cases collected in this study have a greater accuracy, consistency, and validity than any that have been collected on such an extensive scale heretofore." "The committee recognized that it would be very difficult to conduct rigidly controlled studies in the United States during 1953" (p. 3). ". . . the committee recommended four approaches to the problem:

"1. Descriptive epidemiologic studies for each of the areas where mass use of gamma globulin was employed.

"2. A comparison of the severity of paralysis of patients developing the disease immediately before mass use with the severity of those acquiring the disease after receiving gamma globulin.

"3. Study of the severity of paralysis among multiple-case households; . . .

"4. The documentation of administrative aspects of the distribution of gamma globulin" (p. 3).

Appendix B gives reports of epidemiological investigations in thirteen mass inoculation areas, 1953. An evaluation was based on: (1) asymmetry of epidemic curves; (2) shift in age distribution to older groups not receiving gamma globulin, this shift beginning after mass distribution; (3) modification in the duration of epidemics; and (4) differential attack rates. This evaluation turned out to be inconclusive for various technical reasons (pp. 10–18), and in any case was not very encouraging.

The study of severity of paralysis in inoculated and uninoculated patients concluded (p. 21) ". . . its preventive effect in community prophylaxis as practiced during 1953 has not been demonstrated. Also, no modification of the severity of paralysis by gamma globulin was shown. Nevertheless, the committee cannot say that the use of gamma globulin by mass inoculation produced no effect." The need for a more carefully controlled experiment is described.

The multiple-case household study was regarded as adequate for reliable conclusions (p. 85): "They indicate that with the preparations employed and in the dosages used, the administration of gamma globulin to familial associates of patients with poliomyelitis had no significant influence on:

"1. The severity of paralysis developing in subsequent cases.

"2. The proportion of nonparalytic poliomyelitis among the subsequent cases who received gamma globulin before onset.

"3. The classical pattern of familial aggregation of cases in the country at large."

The study of administrative problems may be of value in future work.

Dr. Hammon comments on the study, appropriately reminding the reader of numerous limitations, including the lack of suitable controls. He feels that the modification issue has not been settled, and that the gamma globulin was given too late, but states that the "agent has an extremely limited application in the field of preventive medicine and will not produce dramatic results in general use" (p. 90).

F.M.

RANDOM DIGITS (20,876–21,875)

With this issue, the *Journal* will discontinue publication of random digits. The complete set from which these have been taken is now being published for The Rand Corporation by The Free Press (Glencoe, Illinois) under the title *A Million Random Digits*.

89024	32054	46997	92652	28363
93573	95502	33790	92973	27766
96035	18795	48080	59666	30241
19264	29229	61369	08309	39383
69801	37145	79189	55897	57793
53514	21632	42301	23696	72641
98540	23040	65782	23712	13414
86345	36795	38292	03852	06384
01363	06683	83891	88991	16533
61889	21693	12956	21804	46558
56310	93326	53264	41376	17783
10758	27270	45911	92453	25370
84421	36460	21262	59718	77821
09197	81502	38053	60780	54166
37682	68322	12475	40193	59415
64756	39278	51445	61132	03305
86602	48574	77854	77376	85469
14036	31033	63253	80257	04381
14255	33120	87651	82441	04636
04769	75920	98636	95895	16083
90334	00187	91659	79183	02019
08164	05584	36623	32547	61490
06501	08924	35514	28884	82573
48215	06821	03385	97978	01110
58499	17176	55993	09019	49786
96226	27167	68245	53109	39447
82590	52411	54783	29447	36263
62154	78291	33728	39102	81683
77108	56521	78610	08254	06989
47279	38471	20379	54704	92818
73087	17262	94735	04952	27935
38485	30594	56278	47395	72762
67874	78014	88381	04045	41494
07525	97908	61178	84635	02199
54782	58692	28332	41851	28198
15079	71230	34141	85002	44332
14613	98986	90945	45209	85439
42055	36113	53923	60824	84935
49817	42765	60660	13859	10792
75102	67649	73775	99247	48414

INDEX TO VOLUME 49, 1954

ARTICLES

BOOK REVIEWS

LIST OF REVIEWERS